哼唱检索中一种新颖有效的哼唱信息处理方法*)

马志欣　周利华
(西安电子科技大学多媒体研究所　西安 710071)

摘　要　在基于哼唱的音乐检索系统的研究中,对哼唱输入与数据库进行合理的近似匹配以及有效的检索方法的研究不断深入,但对于哼唱音频信息的有效处理以提取准确有效的旋律特征信息构造查询的研究还不充分。本文在已有的哼唱信息处理方法之上,提出了一种结合了哼唱语音信号增强技术以及时域与频域处理技术的哼唱转谱方法,包括采用了分级音符分割方法、基于规则的音高跟踪方法,并提出一种合理的旋律特征表达的中间格式,用于哼唱查询构造。实验结果证明了这种对哼唱信息转谱处理新方法的有效性和准确性。通过降低哼唱转谱过程中引入的误差,进而可以有效地提高整个音乐检索系统的性能。

关键词　哼唱信息处理,音符分割,音高跟踪,旋律表示

An Effective and Novel Humming Query Processing Method for QBH System

MA Zhi-Xin　ZHOU Li-Hua
(Multimedia Technology Institute of Xidian University, Xi'an 710071)

Abstract　More research in query-by-humming system has focused on the approximate matching approaches to search the database using sung queries. But the researches of effective humming query processing methods and exact melodic character extraction and representation have not been examined maturely. In this paper, an effective and novel humming query processing method has been presented, which combines the speech enhancement technic with time-domain and frequency-domain analysis to do the humming transcription. A hierarchical note segmentation method and a rule-based pitch tracking method have been taken and an intermediate format for melody representation has been proposed. Experimental results show the effectiveness and accuracy of the humming transcription, thus the performance of following melody based searching can be improved.

Keywords　Humming query processing, Note segmentation, Pitch tracking, Melody presentation

1　引言

基于内容的多媒体信息检索技术成为解决如何在信息网络中自然、方便并且迅速有效地找到自己需要的多媒体信息的重要方式。

对基于内容的音乐信息检索(Music Information Retrieval, MIR)技术的研究在近 10 年来受到越来越多的关注。基于内容的音乐信息检索技术是一个涉及交叉学科的研究方向,研究者包括音乐家、计算机科学家、图书馆及信息科学家、工程师、认知科学家、音乐心理学家以及其它各种专业人员。研究的目的是努力找到创新的基于内容的检索模式。

基于内容的音乐检索涉及音乐旋律表达问题、音乐旋律特征提取问题、用户查询构造问题、音乐旋律匹配问题以及音乐数据库构造问题等很多方面,这些都是建立一个完整、有效的音乐检索系统的关键。

2　哼唱检索

在音乐信息检索系统中,用哼唱构造查询(Query- By-Humming, QBH)的方法提供了一种自然和方便的用户查询接口而得到了广泛和深入的研究。哼唱音乐检索技术最早提出于 20 世纪 90 年代,早期研究有文[1~3]等实验系统。

哼唱检索的用户查询构造的过程,就是从用户的哼唱音频输入得到用户所要表达的音乐旋律信息,即通过一系列的信号和数据处理过程将原始音频信息转换成为符号表示的音乐旋律信息,从而可以被计算机有效地用来作为检索系统的查询输入数据。

对哼唱信息的获取和处理有多种方法,处理的主要目的是提取出其中表达的音乐旋律特征。对信号的处理过程使用时域方法、频域方法或时域频域相结合的处理方法进行音符分割、音高跟踪,得到旋律特征数据序列,之后通过特定的组织方式将旋律特征序列作为查询输入,在数据库中进行旋律特征匹配,找到检索目标。

本文的工作面向哼唱检索的最前端:哼唱信号处理。即对哼唱音频信号进行处理,获取旋律特征信息,并将旋律信息表示为一种合理的中间格式,可以直接或变换后用于不同的音乐检索系统,进行查询构造。

3　哼唱信息处理方法

对哼唱音频信号的处理方法可以采用一般对语音信号分析处理的方法,但又有所不同:语音信号处理针对目的的不同,处理的重点与具体采用的方法也不同。

我们对哼唱信息处理的过程分为 3 个部分:预处理,音符

*)本文得到国家部委科技电子预研项目资助(413160501)。马志欣　博士研究生,研究方向:网络多媒体、多媒体信息检索技术;周利华　教授,研究方向:多媒体技术、网络信息安全技术。

估计、旋律表达。这几个部分中，准确进行音符估计是系统的难点所在。系统结构见图 1。

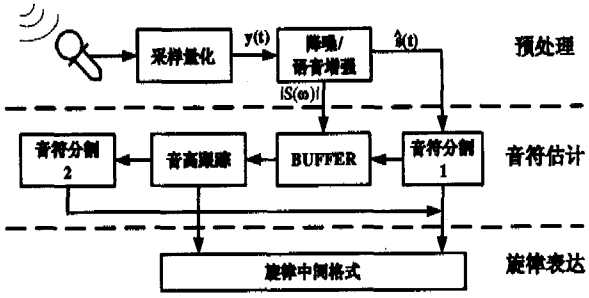


图 1　系统框图

3.1　预处理

预处理部分包括声音录制、采样量化、降噪等操作。一般使用声卡和麦克风进行录音采样来得到数字化的哼唱音频数据。不同的系统采样参数不同，但差别不大。人的发声基频范围在 80Hz～800Hz 以内，综合考虑系统所需的数据精度和计算复杂度，确定了预处理程序音频采集格式为 11025Hz/8bit/mono，数字化后去除信号直流分量。

现有哼唱检索系统大多对数字化的音频信息只做简单低通滤波甚至不进行加工，还有些系统采集哼唱输入时要求在安静的房间使用专业麦克风，只有少数文献提到对音频的预处理[4]。在实际应用环境中，一般面对的条件是配置声卡的个人电脑和普通的麦克风，且无法避免环境噪声的干扰。所以我们要建立一个强健的系统，必须强调对输入信号采用降噪方法，或者说语音增强的方法。

对语音信号中加性噪声的抑制方法有多种。对于哼唱信息，我们处理的主要目的是为了有利于后续的音符分割和音高跟踪，所以采用减谱法[5]，可以简单而有效地对音频信号的加性噪声进行降噪处理。我们采用了一种改进的减谱法，其效果优于传统方法，如图 2 所示。

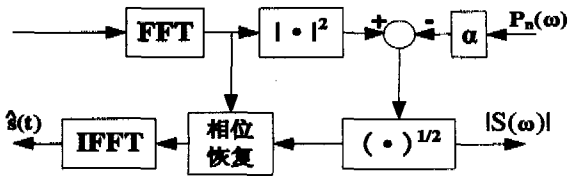


图 2　改进的减谱法降噪/语音增强

用 $P_y(\omega)$，$P_s(\omega)$ 和 $P_n(\omega)$ 分别表示带噪语音 $y(t)$、增强后的语音 $\hat{s}(t)$ 和加性噪声 $n(t)$ 的功率谱，改进的减谱法公式如下：

$$P_s(\omega)=\begin{cases}P_y(\omega)-\alpha P_n(\omega) & ,P_y(\omega)\geqslant\alpha P_n(\omega)\\ 0 & ,P_y(\omega)<\alpha P_n(\omega)\end{cases} \quad (1)$$

取 $\alpha=5$（$\alpha=1$ 时即为传统减谱法）效果较好。

其中的噪声功率谱，我们采用以下方法估计而得：系统软件使用模拟指示灯和声音提示，用户选择开始录音时，系统类似电话留言机方式提示用户，在看到指示灯由红变绿并发出“滴”声后开始哼唱。从用户选择开始录音到系统指示灯变成绿色并发出提示音之间经历 1s 时间，系统在这 1s 并非等待，而是录制 1s 的环境噪声信号，对其进行谱分析，得到 $P_n(\omega)$ 的估值。

图 2 给出了降噪效果图。图 2 上图为原始带噪信号，下图为降噪后的信号。可以看到信号最前端录制的 1s 噪音信号段。

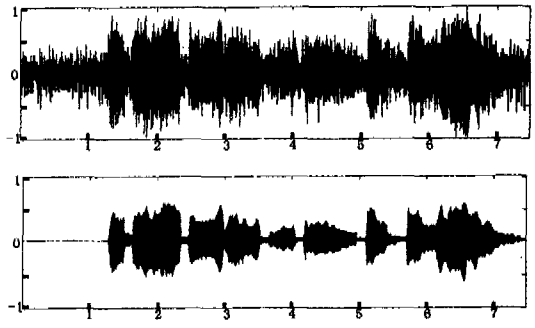


图 3　降噪效果图

3.2　音符估计

音符估计部分包括两个重要方面：音符分割和音高跟踪（基音检测）。音符分割从哼唱音频信息中找到每个独立音符的起止时间信息，将音频按音符的变化进行分段，得到每个独立音符的位置特征。音高跟踪从表达每个音符的音频信息中获取其基频信息，也就是音高特征信息。

3.2.1　音符分割

在一般语音处理的过程中，需要进行端点检测，以确定语音信号中的各种段落开始和结束的位置，如判断出哪些部分是语音信号，哪些部分是静默或噪声信号，进一步要判断音素、音节、词等。对于哼唱音符分割的处理相对简化，可以采用语音识别端点检测类似方法。常用的端点检测方法有：基于能量、基于能量与过零率、基于信息熵、基于频域特征等。如何准确地进行哼唱音符分割，是哼唱检索系统中最困难的问题[6]。

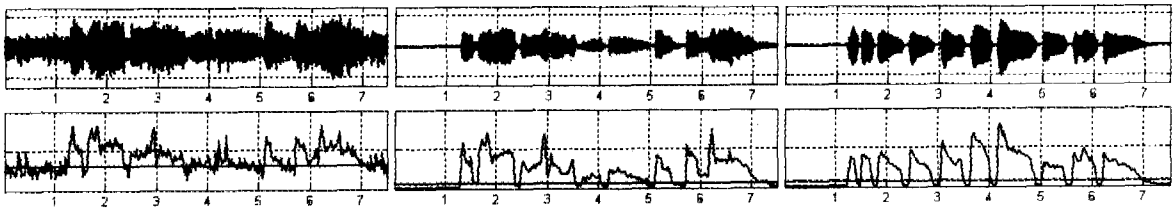
我们采用了分级的音符分割方法，使用时域方法与频域方法相结合，既简化计算过程，又可以比较有效而准确地进行音符分割。

首先采用基于能量跟踪的方法作为第一级：由于有效地进行了降噪处理，信号有较高的信噪比。这种情况下，通过计算信号的短时能量，就可以通过能量跟踪找到每个音符语音段的开始及结束时间。但对于少数连唱的音符无法分割开来。

现有不同的哼唱检索系统对哼唱输入的要求也不同，大致可以分成两种：①要求用户采用“Da”或“Ta”等爆破音加浊音音节的方式哼唱；②对哼唱方式不做要求，可以使用任意发音哼唱，比如使用歌词演唱或歌唱哼唱混合等。

前一种方法充分利用哼唱检索音乐的特点，好处显而易见。由于唱出每个音之前都会自然形成短暂的停顿，而且哼唱辅音相同，每个音符的谱特征相似，哼唱得到的音频信息中每个音符的音量幅度变化也不大，这样为音符的分割带来了好处，可以方便地使用能量跟踪方法进行较准确的音符分割，对后续的音高跟踪也有好处。图 4 给出了使用能量跟踪对带噪哼唱、降噪用歌词哼唱以及降噪用“Da”声哼唱片段进行处理的情况，可以很清楚地看到对于带噪信息很难使用能量跟踪准确分割音符；对于用歌词哼唱的音频，较难估计和确定最佳阈值，音符分割误差较大；可以看到使用“Da”声哼唱的明显优势。

有了音符分割的数据，音高跟踪的工作也得以简化并更加有效。第二级的音符分割基于音高跟踪的结果得到，目的是把第一级没有分割开的连唱音符根据音高显著变化特征，即基音频率不同而进一步分割开来，参见图 6。



(a)带噪哼唱　　(b) 降噪用歌词哼唱　　(c) 降噪用“Da”声哼唱

图 4　能量跟踪

3.2.2　音高跟踪

在图 1 中可以看到，语音增强处理过程得到的增强后的语音谱数据存放在“BUFFER”之中，有两个用途：一是用于 IFFT 得到时域的语音增强信号，进行音符分割；另一个就是用于后续进行基音检测。我们采用谐波积频谱 HPS(Harmonic Product Spectrum)方法[7,8]进行基间检测，直接使用“BUFFER”中数据，无需再次进行语音的谱计算。HPS 根据 N 次谐波峰值关系得到基频，如图 5 所示。

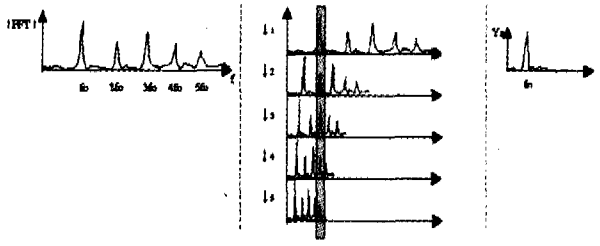


图 5　HPS 示意图

HPS 所得到的音高分辨率与 FFT 的点数相关，音高是采样频率/FFT 点数的倍数。若需要高解析度，可以将音帧补 0 增加点数，提高频率分辨率。另外需要考虑的是，HPS 得到的音高容易受到共振峰的影响。

我们采用了基于规则的基音检测过程，制定了如下规则：

① 基音频率范围在 80～800Hz 以内；

② 独立音符的哼唱时长大于 100ms；

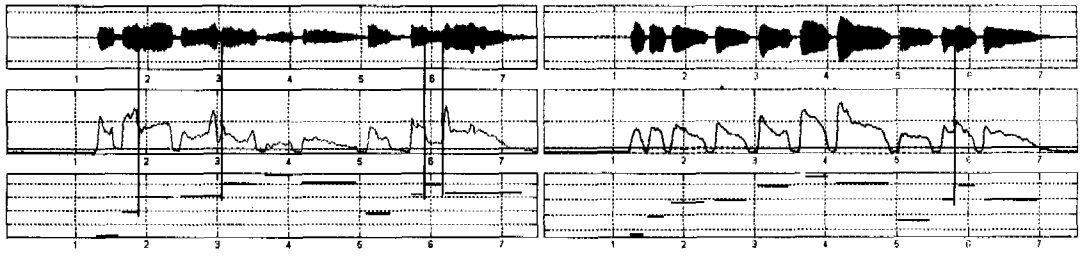
③ 相邻音符的哼唱音高差大于 50 音分；

④ 音符中间部分的哼唱音高准确度大于开始和结束部分。

根据规则①，在数据存入 BUFFER 时进行等效带通滤波操作，仅处理 80～3200Hz 数据(包含 4 个以上谐波的范围)，这样既保留了有效信息，又节省存贮空间和处理时间。根据第一级音符分割得到的音符起始点时间信息，对表示音符的有效帧应用 HPS 得到每帧的基音频率。

我们对 HPS 算法进行改进，以减少共振峰的影响。在进行 HPS 之前，以频谱最高峰值-35dB 作为阈值检测得到最低频率的峰值，作为 HPS 检测峰值的频率下限。检测 HPS 的次高峰值，如果次高峰值点频率约为最高峰值点频率的 1/2，则将次高峰值点频率作为基音频率。这样可以有效地找到真实基音频率。

我们对声音信号进行的谱分析处理采用每帧 512 点，重叠 374 点分帧的汉明窗。为了在最低频率处的频率分辨率基本达到一个半音，对数据补 0 到 2048 点，实验表明在目前的一般普通个人电脑上的计算效率完全可以满足要求。这样得到的初始帧约对应为 12.5 毫秒，处理野点后，两两求均值合并，这样音高估计的最小时间单位约 25ms，作为最小音高帧。根据规则②和规则③对被第一级音符分割出的连续时间信号进行音高跟踪，得到是否有不同音高的音符粘连的信息，作为第二级音符分割信息。根据规则④对独立音符的所有数据帧音高值加权平均作为音符的最终音高频率估计值。通过音符分割、音高跟踪得到的音高轮廓如图 6 所示，可以看到竖黑实线指示的从音高跟踪得到的第二级音符分割信息。



(a) 用歌词哼唱　　(b)用“Da”哼唱

图 6　音高跟踪和第二级音符分割

这样，我们就由两级音符分割得到了每个独立音符的起始和结束时间(以帧序号表示)，由音高跟踪得到了每个音符的基音频率值。

3.3　旋律表达

得到音符序列的音高和时间特征数据后，按照一定的组织结构将这些数据表示成中间格式，进行旋律表达，供后续检索系统构造查询。我们设计旋律表达中间格式的指导思想是尽量保持特征数据的原本面貌，且用于后续检索系统时无需复杂的数据转换。根据人类对音乐的认知和旋律记忆方法研究结果，我们知道旋律轮廓是人类旋律记忆最重要的部分，即人类记忆和回放的是比较准确的旋律轮廓信息，也就是比较准确的相对关系的音高和音长信息。因此，大多数音乐信息检索系统所采用的查询构造方法，其音高特征多数采用相对音高特征信息，音长特征一般采用相邻有效音符的起始时间

距离 IOI(inter-onset interval)作为音长特征，只是不同系统的量化和组织方法有差别。由于后续检索系统的具体处理方法各有不同，故我们对原始数据不进行量化。根据以上分析所设计的旋律特征中间格式采用如下步骤得到：

① 设音符的音高频率序列为(f_0，f_1，…，f_m)，使用相邻音符的音高频率 f_{i-1} 和 f_i 计算得到相邻音符的音高差数据，以音分为单位得到音高差特征序列为(Δ_{p_0}，Δ_{p_1}，…，Δ_{p_m})，其中：

$$\Delta p_i=\begin{cases} * & ,i=0 \\ 1200\times\log_2(f_i/f_{i-1}) & ,0<i\leqslant m \end{cases} \tag{2}$$

② 设音符的起始帧序号序列为(t_0，t_1，…，t_m)，使用相邻音符的起始帧序号 t_{i-1} 和 t_i 计算得到音长特征序列为(Δt_0，Δt_1，…，Δt_m)，其中：

$$\Delta t_i=\begin{cases} * & ,i=0 \\ t_i-t_{i-1} & ,0<i\leqslant m \end{cases} \tag{3}$$

③ 用音高差特征序列和音长特征序列构成旋律表达中间格式：〈ΔP，ΔT〉。

4 实验结果

我们通过实验来评估上述方法的性能，实验的目的是验证所采用方法的有效性并评估其性能。选择了未经专业音乐训练的男女各 3 人参与了实验，每人分别用歌词和"Da"声哼唱选定的 5 首流行歌曲的片断，每段录音均不超过 1s+10s，使用普通带声卡 PC 连接通麦克风，在有环境噪声的实验室里进行录音。这样，一共得到 60 个哼唱录音片断。按照本文所提出的方法编制哼唱信息处理测试程序，以文本文件方式给出处理结果。

所有录音由专业人士通过专业音频处理软件人工进行音符分割以及音高估计，作为评估参照标准。因为评估目标仅为哼唱信息处理方法本身，故人工处理完全以具体哼唱内容为准，不参照曲谱，评估结果只体现本文方法与人工处理的差别。实验分别评价音符分割和音高跟踪性能。人工处理对哼唱片断中的哼唱滑音均按照不同音符处理。人工处理得到的数据共包含 1090 个独立音符，歌词哼唱包含 559 个音符，"Da"声哼唱包含 531 个音符。

音符分割评估标准：设人工标定音符分割点总数为 S_t，在误差范围位置内，测试程序正确找到的分割点数为 S_r，在没有分割点的地方多找出的分割点数为 S_e，定义音符分割准确度 R_S 为：

$$R_S=(S_r-S_e)/S_t\times100\% \tag{4}$$

音高跟踪评估标准：设人工标定音符音高差总数为 P_t($P_t=S_t$)，因音符分隔错误而无法对应的音高差数为 P_s，可对应比较的音符音高差与参照标准误差在±40 音分范围内被认为是准确的音高差总数 P_r，定义音高跟踪准确度 R_P 为：

$$R_P=P_r/(P_t-P_e)\times100\% \tag{5}$$

定义音符估计综合准确度 R_T 为：

$$R_T=P_r/R_t\times100\% \tag{6}$$

实验结果如表 1～4。

表 1　针对不同唱法的准确度

	R_S	R_P	R_T
使用歌词哼唱	92%	96%	82%
使用"Da"声哼唱	99%	99%	96%

表 2　针对不同实验者的准确度

	R_S	R_P	R_T
男 1	96%	97%	91%
男 2	91%	97%	80%
男 3	95%	97%	88%
女 1	98%	99%	97%
女 2	97%	97%	91%
女 3	95%	95%	86%

表 3　针对不同旋律的准确度

	R_S	R_P	R_T
歌曲 1	98%	98%	95%
歌曲 2	97%	96%	92%
歌曲 3	92%	96%	82%
歌曲 4	95%	97%	88%
歌曲 5	94%	99%	88%

表 4　综合准确度

	R_S	R_P	R_T
全部哼唱	95%	97%	89%

从实验结果我们可以直观地看到以下几个规律：

① "Da"声哼唱方法的音符分割、音高跟踪和音符估计综合准确度均高于使用歌词哼唱方法；

② 不同演唱者之间比较，音符分割准确度差异大，音高跟踪准确度差异小；

③ 不同歌曲旋律，复杂度不同，影响哼唱效果；

④ 整体音符分割准确度低于音高跟踪准确度。

分析：

使用"Da"声哼唱的方法，其音符分割准确度高是显而易见的。因为这种哼唱方法人为地避免了多数连唱音，尤其是相同音高的连唱音，使用音高跟踪是难以分割开的，所以，"Da"声哼唱方法的音符分割准确度远大于歌词哼唱方法。在使用歌词哼唱时，歌词发音中包含清音发音，在演唱时如果音长较短，一般演唱者会使用清音发音，造成音高信息误差。而使用"Da"声哼唱时，所有发音都是浊音发出，所以其音高跟踪的准确度也比较高，当然音符估计综合准确度也一样较高。

对相同旋律的表达，不同的人歌唱能力不同，连音、滑音等演唱习惯对音符分割的影响较大。当然，唱歌跑调的误差并不是本文方法所讨论的内容。

歌曲旋律的差异导致哼唱效果的差异，这一点仍然和歌唱能力有关。因为参与实验者均为未经专业音乐训练的人，所以在对不同难度旋律的哼唱时表现出来的效果差异比较明显。

结论　从整个方法系统的设计实现和实验结果可以看到，本文提出的哼唱信息处理方法有效地进行了从哼唱音频到音符序列的转谱处理。在已有的哼唱信息处理方法之上，所实施的结合语音信号增强技术和时域与频域处理方法有效地去除了加性噪声的影响，明显降低了音符估计的处理难度。分级音符分割方法，进一步提高了音符分割的精度，使得整个哼唱转谱的整体准确度较高。通过实验样本得到的综合准确度达到了 89%。

实验结果充分证明了这种新的哼唱信息处理方法是准确有效的。通过哼唱信息处理有效地降低了哼唱转谱过程中引

（下转第 190 页）

试样本被判属为第 l 类。

(6) 若对测试样本是第一次识别，则重新修正各类的样本均值向量。否则，若为第二次识别则转步骤(7)。

$$m_i=\overline{X}_i=\frac{1}{n_i+n_j}(n_i*m_i+\sum X_{ki})$$

$$m=\overline{X}=\frac{1}{2}\left(m+\frac{1}{C}\sum_{i=1}^{C}m_i\right)$$

式中 m_i 是第 i 类模式的类均值向量($i=1,2,\cdots,c$)，n_i 是第 i 类模式训练样本个数，n_j 是被判属为第 i 类模式的测试样本数。转步骤(3)。

(7)作正确率测试，并将识别正确率作为最终结果输出。

该算法利用测试样本对各类样本均值及总体样本均值进行了修正，并利用修正后的样本均值重新计算类间及类内协方差矩阵，再进一步求出 Fisher 最佳鉴别矢量集。利用该鉴别矢量集进行特征提取时，其效率应明显高于原鉴别矢量集的提取效率。

4 实验及分析

本文采用 ORL 人脸库进行对比实验，该库由 40 人的人脸图像组成，每人 10 幅。原图像为 92×112 像素，先对所有的训练及测试样本进行两次小波变换，将图像变换为 23×28 像素。

实验以每人的前 5 幅图像作为训练样本，后 5 幅作为测试样本。因此训练和测试样本均为 200 幅。首先，采用基于类间离散矩阵 S_b 的离散 K-L 变换，对原始特征进行特征提取，由于模式的类别数为 40，经特征提取后模式特征的维数降为 39 维。然后，模式分类采用最小距离分类器。对于距离的度量分别采用欧氏距离和绝对值距离。

表 1 为采用三种特征提取算法的识别率。

表 1 识别正确率的百分比

	欧氏距离	绝对值距离
PCA	89.5	90
Fisherface	93	92
DA-Fisherface	94	94

从实验结果可以看出，DA-fisherface 方法的模式识别率比 PCA 及 Fisherface 方法都有显著的提高，这是由于对小样本数据而言，使用训练样本均值作为各类别均值及总体均值的估计，会产生较大的偏差，进而引起样本协方差矩阵的偏差，并导致模式识别率的下降。而 DA-fisherface 方法利用测试样本数据对上述偏差进行了修正，因此，其识别的正确率有明显提高。

结论 本文提出的具有动态调整功能的 Fisherface(DA-Fisherface)方法，是利用测试样本实现对类别均值的修正，进而对由于小样本数据所造成的类间及类内协方差矩阵的偏差进行了修正，并进一步实现了对 Fisher 鉴别矢量集的优化。实验结果证明了该方法的有效性。

由于考虑特征提取的效率问题，本文没有对总体协方差矩阵进行修正，若对总体协方差矩阵进行修正，则模式分类的正确率有望得到进一步提高。

参 考 文 献

1 Belhumeur P N, et al. Eigenfaces vs. Fisherfaces: Recognition using class specificlinear projection. IEEE Trans. Pattern Anal Machine Intell, 1997, 19(7): 711～720

2 Hong Z Q, Yang J Y, et al. Optimal discriminant plane for a small number of samples and design method of classifier on the plane. Pattern Recognition, 1991, 24(4): 317～324

3 Liu K, Yang J-Y, et al. An efficient algorithm for Foley-Sammon optimal set of discriminant vectors by algebraic method, International Journal of Pattern Recognition and Artificial Intelligence, 1992, 6(5): 8817～829

4 Hua Yu, Jie Yang. A direct LDA algorithm forhigh-dimensional data—with application to face recognition. Pattern Recognition, 2001, 34(11): 2067～2070

5 Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer-Verlag, 1995

6 Mika S, Ratsch G, Weston J, Scholkopf B, Muller K. Fisher Discriminat Analysis with Kernels. In: Proc. of the IEEE Neural Networks for Signal Processing Workshop, Madison, 1999. 41～48

7 Scholkopf Mika S, et al. Input Space Versus FeatureSpace in Kernel-Based Methods. IEEE Trans on Neural Networks, 1999, 10(5): 1000～1017

8 Thomaz C E, Gillies D F, Feitosa R Q. A New Covariance Estimate for Bayesian Classifiers in Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, FEBRUARY, 2004, 14(2): 214～223

9 Wang Xiaogang, Tang Xiaoou. Dual-Space Linear Discriminant Analysis for Face Recognition Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition

10 Wang Xiaogang, Tang Xiaoou. Using Random Subspace to Combine Multiple Features for Face Recognition Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition

(上接第 172 页)

人的误差，进而提高音乐检索系统的性能。实验表明，这种方法尤其适合于使用类似“Da”声哼唱的系统。

用于哼唱查询构造的合理的特征信息表示中间格式，既可以避免后续系统进行复杂的运算，又提供了完整的特征信息，可以方便有效地接入到不同的音乐检索系统中。

进一步的研究工作着眼于对用户使用任意声音自由哼唱旋律的方法结合语音识别领域的方法和成果来进行有效的识别处理，而处理方法的重点和难点仍然在于音符分割方法的研究。

参 考 文 献

1 Kageyama T, Mochizuki K, Takashima Y. Melody retrieval with humming. In: Proc. ICMC1993. Tokyo: ICMA, 1993. 349～351

2 Ghias A, Logan J, Chamberlin D, et al. Query By Humming Musical Information Retrieval in An Audio Database. In: Proc. ACM Multimedia 95. San Francisco: ACM press, 1995. 231～236

3 McNab R J, Smith L A, Witten I H, et al. Toward the digital music library: tune retrieval from acoustic input. In: Proc. ACM Digital Libraries. Bethesda: ACM press, 1996. 11～18

4 Pollastri E. A pitch tracking system dedicated to process singing voice for music retrieval. In: Proc. of ICME2002. Switzerland: IEEE, 2002. 341～344

5 赵力. 语音信号处理. 北京：机械工业出版社，2003

6 Meek C, Birmingham W. Johnny cant't sing: A comprehensive error model for sung music queries. In: Proc. of ISMIR2002. Paris: 2002

7 拉宾纳 LR, 谢弗 RW. 语音信号数字处理. 朱雪龙，等译. 北京：科学出版社，1983

8 Noll M. Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate. In: Proc. of the Symposium on Computer Processing Communications. New York: Polytechnic Press, 1970. 779～797