

基于 OMR 技术的乐谱图像匹配算法

陈根方¹, 张文俊²

(1. 杭州师范学院计算机系, 杭州 310036; 2. 上海大学影视艺术技术学院, 上海 200072)

摘 要: 提出了一种数字乐谱图像识别与匹配方法, 采用光学乐谱识别(OMR)技术识别数字乐谱图像建立对应的 MIDI 文件, 利用一一对应算法进行匹配, 根据相似度来判断同一音乐作品的不同数字版本的乐谱图像, 并在理论上就此算法的有效性进行了论证。同时, 仿真实验结果表明这一方法能有效地实现对不同数字版本的同一音乐作品的乐谱图像的识别、归类与校正。

关键词: 匹配算法; 乐谱图像; 光学乐谱识别(OMR)

Music Image Matching Algorithm Based on OMR Technique

CHEN Genfang¹, ZHANG Wenjun²

(1. Department of Computer, Hangzhou Normal College, Hangzhou 310036;

2. School of Film & TV Arts and Technology, Shanghai University, Shanghai 200072)

【Abstract】 This paper proposes a recognition and matching algorithm. It gets two MIDI files with OMR technique from the two musical score images, and then using the algorithm, it gets the degrees of similarity, with which it can judge whether the two musical score images are the same musical works. It provides the proof of the validity of the algorithm.

【Key words】 Matching algorithm; Music image, Optical music recognition(OMR)

随着数字媒体技术的发展, 数字化音乐得到迅速普及, 已经成为现代音乐创作、传播、存储等的主要形式。纸质媒介作为传统的主要音乐载体, 已经逐渐被数字载体所取代, Internet 也已成为音乐的主要交流与传播平台。光学乐谱识别(Optical Music Recognition, OMR)是利用扫描仪、数码相机等数字设备把纸质媒介上的信息转化为数字图像信息, 音乐乐谱被转化为数字乐谱图像, 再利用专用的图像处理软件, 识别出数字乐谱图像中的音乐信息, 并以流行的格式(如 MIDI 格式)存到存储器中。目前已有商业化的 OMR 软件系统, 如 Neuratron Ltd. 的 PhotoScore, 其识别率可达 90% 以上^[1]。数字乐谱图像存储在数据中心或在 Internet 中流动, 这是一个巨大的音乐资源, 这些乐谱图像能被传承、复制, 打印, 演奏, 出版等。由于音乐作品的出版量大、版本与形式不同, 以及不同的数字化过程与实现方式, 经数字化后会造成本一音乐作品的不同的数字乐谱图像, 比如, 网络上(或数字图书馆)存在着大量数字乐谱图像, 而其中有些数字乐谱图像可能是同一音乐作品的不同数字版本, 这对音乐作品的使用与交流带来诸多的不便。因此, 需要研究和建立数字乐谱图像识别、归类 and 校正的方法。

本文提出了一种基于 OMR 技术的数字乐谱图像匹配算法, 利用 OMR 识别数字乐谱图像建立对应的 MIDI 文件, 依据承载的音乐信息, 采用一一对应算法进行模式匹配得到相似度, 以实现同一音乐作品的不同数字版本的判别与归类。

1 数字乐谱图像的识别与匹配方案

数字乐谱图像的识别与匹配方案的框架如图 1 所示。具体实现步骤是:

(1) 先根据音乐领域知识(主要是记谱法)和数字乐谱图像的特点, 对数字乐谱图像进行预处理和 OMR 识别。OMR 技术利用了数字乐谱图像的一些重要特征和记谱法的一些相关知

识, OMR 软件系统都是基于知识的软件系统。大多数的 OMR 技术都采用先识别谱线, 然后识别音符、时值, 最后识别音乐表情符号的处理与识别顺序, 其中有些识别技术要求谱线的水平倾斜度小于 $\pm 10^\circ$ ^[2]。利用数字图像处理技术对要比较匹配的两个图像(分别设为 MI1 与 MI2)进行预处理, 如图像倾斜度校正和去噪声等。再利用 OMR 技术对 2 个乐谱图像 MI1 和 MI2 分别进行音乐信息识别。

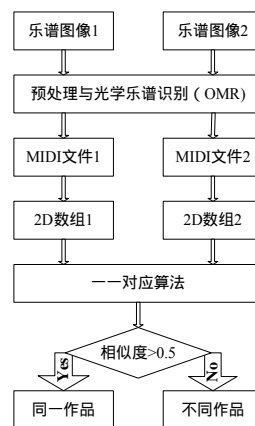


图 1 乐谱图像匹配算法流程

(2) 依据识别出来的音乐信息, 以 MIDI 格式予以存储, 建立 2 个 MIDI 文件(分别标记为 MIDI1 和 MIDI2)。

(3) 根据 MIDI 格式要求, 把 MIDI 中的音乐信息转化为 2 个二维数组(分别记为 Array1 和 Array2), 这 2 个数组中只包

基金项目: 2005 浙江省教育厅计划课题(0561XP51)

作者简介: 陈根方(1968 -), 男, 讲师、博士生, 主研方向: 算法, 数字乐谱图像识别, 数字音乐; 张文俊, 教授、博导

收稿日期: 2006-10-31 **E-mail:** cgf@hztc.edu.cn

含 MIDI 文件中的音符的音高与时值这 2 个信息，一维记录和声，另一维记录时值。比如对于独奏的钢琴乐谱，同时发声的音符最多不超过 10 个，所以记录和声的这一维的下标最大为 10，但是另一维由于是时间维，因此根据乐曲的长短，有所不同。

(4)利用一一对应算法来比较 2 个数组 Array1 和 Array2 横向的相似程度，得到相似度(记为 S_{ome})。

(5)由于不同的 2 个作品的相似度很低，平均记为 $S < 0.5$ 。这是一个相对固定的值。

(6)当 $S_{ome} > 0.5$ 时，可认为这 2 个乐谱图像是同一个音乐作品。否则认为这 2 个乐谱图像不是同一个音乐作品。

(7)如果是同一个音乐作品的不同乐谱图像，可以利用 2 个乐谱图像所显示的音乐信息，分别对其中的一些缺陷进行校正或修复等工作。

2 图像相似度度量算法

音乐作品是由不同的音符所构成的有限长的有序序列，音乐的演奏是二维音响的有序的演绎，一维是时间维，另一维是瞬时时间内不同音高、音色的重叠共响，记谱法利用五线谱和连谱号作为框架记录二维的音响演绎，2 个或以上的音符可以同时发声，不同的音符按五线谱中的记录顺序一一发声，形成丰富多彩的音乐形象。

如果任意改变一音乐作品中的 2 个音符之间的前后关系，就可能成为另一音乐作品，即音乐作品是有序的音响构成的有限长的序列。可以用一个二维数组来记录音乐作品的音响演绎，如果去掉音乐表情记号与演奏记号，则记谱法就是二维数组在纸质媒介上的最好表现形式之一，在下面讨论的二维数组中，一维是时间维，一维是记录遵循 MIDI1.0 标准的不同音高的整数存储空间。

具体算法如下，设这 2 个二维的有序序列是 2 个二维数组：

(1)记 2 个二维数组为 $A[m][n]$ 与 $B[s][t]$ ，不妨设 $m \geq s$ ；

(2)用符号 $X[i]$ 表示一维数组 $A[i][1]$ ， $1 \leq i \leq m$ ，同样，用符号 $Y[j]$ 表示一维数组 $B[j][1]$ ， $1 \leq j \leq s$ ；

(3)对于 $k(1 \leq k \leq m-s)$ ， $C[k]$ 的定义为：初始时 $C[k]=0$ ，对于每一个 $r(1 \leq r \leq s)$ ，若 $X[k+r]=Y[r]$ ，则 $C[k]=C[k]+1$ ；

(4) 2 个二维数组 A 与 B 的相似度 S 记为：

$$S = \max(C[k]), 1 \leq k \leq m-s$$

其中步骤(3)利用了一一对应的方法进行的相似度计算，可利用 LCS 算法^[3]来定义另一不同的相似度。

3 算法有效性论证

由于音乐是二维音响的演绎，记谱法能很好地记录音乐信息，它的横向是时间维，由连谱线连起来的纵向构成另一维，不妨称之为音高维。这一维常是和弦或不同乐器的音高或演奏信息等，在以下的讨论中，音高维中只包括音符的音高高低，不考虑其它音乐信息(如演奏信息、表情信息)，简单地说，由时间维与音高维构成乐谱。

用 N 表示记谱法中能使用的所有音高的集合，并记 N 中不同音高的个数为 $n=|N|$ 。例如，在 MIDI1.0 中，所有的音高分别被表示为 0~127，中央 C 是 60，这样，从 MIDI1.0 的角度来说， N 是由 0~127 共 128($=n$)个整数所构成。

显然，音高维中的每个音高 $p \in N$ ，这样的不同的 p 共可以有 n 个，但是，常用的音高或者说使用频率高的音高是比较集中的。如果音乐作品中，音高维中音高的个数不超过 m 个，则可以产生 $C(n,m)$ 种不同音高维的组合。比如在音乐作

品《蓝色多瑙河》(首段)中，音高维中音高的个数不超过 5 个(15 小节第 2 拍和弦是 5 个不同的音高)，而最低音是大字一组的 A，最高音是小字两组的 d，就是说这个作品中可以取来用的不同的音高的个数为 41 个，这 41 个音高包括了调式(它是 D 大调)之外的音高，如果只使用调式之内的音高，这里只有 25 个不同的音高(有一个变化音)。这样，这个作品可以产生的不同音高维组合是 $C(24, 5) = 42 504$ 。这么多变化中如果加上和弦的使用原则等约束条件，则可以使用的不同的音高维可能少得多，但也足以产生这首美妙的音乐了。

在计算机中，为了统一演奏速度，必须要有一个统一的基准时间片，这个基准时间片叫时间戳，记为 t ，其它所有拍子的时间长短都是以这个 t 为单位的，比如，设一拍为 $192t$ ，则二分之一拍就是 $96t$ ，四分之一拍就是 $48t$ ，二拍就是 $384t$ ，等等。如果对于 $1t$ ，计算机的对应的时间是 $0.01s$ ，则一拍对应的的时间就是 $1.92s$ 。时间维的分割就是以时间戳的某个整数倍为单位的。

一个音乐作品常有时值最短的音符，比如在《蓝色多瑙河》(首段)中，时值最短是二分之一拍，如果以 $192t$ 为一拍的话，则二分之一拍就是 $96t$ ，这样的话，记录这首作品的时间维可以以 $96t$ 为分割单位，每过 $96t$ ，就记录一次音高，构成一个音高维序列，无论这个时间上有没有新的音高出现；比如《蓝色多瑙河》(首段)以 $96t$ 为分割单位，可以得到 192 个音高维(事实上就是 32 个小节，每小节 3 拍，1 拍 2 个八分音符为一个分割单位， $32 \times 3 \times 2 = 192$)，而其中的 100 个音高维是空值。

下面来分析一般情况，设音乐作品由 L 个音高维构成，每个音高维由不超过 m 个不同音高组成，作品中所有不同的音高共有 n 个，则可以产生不同的音乐作品数为

$$W = [C(n, m)]^L$$

显然这是一个巨大的数量。当然有一些约束条件，比如：和弦的构成规律、作曲法、配器法等，会使 W 中悦耳的音乐大大减少。简单地说， W 中绝大多数是不悦耳的音乐。

这样，随机产生的 2 个音乐作品(如计算机作曲)，如果它们的 L, m, n 都相同，出现一模一样的可能性为 $1/W^2$ 。对于普通的音乐作品，这是一个非常小的数。

下面用一一对应的算法来讨论 2 个任意作品的相似度，这里只讨论 $m=1$ 的情况，即单声部作品，并假设作品的长度都是 L ，显然有：

长度为 L ，不同音高数为 n 的不同的组合音乐作品数量为

$$T = n^L$$

而不同的 2 个作品的比较，组合比较数为

$$T(T-1)/2$$

可以得到所有不同的 2 个作品一一对应比较时相同的总数为

$$n^{L-1}(n^{L-1}-1) * n * l / 2$$

所以任意 2 个作品(长度为 L ，单声部，所有不同音个数为 n)的相同音符平均个数为

$$average = [n^{L-1}(n^{L-1}-1) * n * l / 2] / T = \frac{L * (n^{L-1}-1)}{n^{L-1}-1} \approx \frac{L}{n}$$

换句话说，2 个任意长度为 L 的单声部作品的一一对应相似度 S 为

$$S = \frac{average}{L} = \frac{1}{n}$$

下面来考察同一作品的不同版本通过某一 OMR 系统后，

得到 2 个 MIDI 文件，由于常见的 OMR 系统的识别率(记为 P)都在 90% 以上^[1]，即 $P > 90\%$ ，2 次识别事件是相互独立的，因此这 2 个 MIDI 文件的一一对应相似度 S_{one} 为

$$S_{one} = P^2$$

由于 $P > 90\%$ ，因此 $S_{one} > 81\%$ ，同时，由于普通的音乐作品的不同音符的个数 $n > 2$ ，因此有

$$S_{one} > S > \frac{1}{2}$$

由此可见，同一作品的不同版本之间的相似度远大于不同作品之间的相似度。

4 仿真实验

在仿真实验中，从大众乐谱网站选取了 52 首不同的作品^[4]，包括钢琴曲、民歌、舞曲、交响曲主题等不同体裁的音乐作品的数字图像乐谱。其中有 3 首是同一作品《嘎达梅林》的不同数字版本。

先就选取的 52 个作品分别利用 OMR 软件系统进行音乐信息的识别与提取，产生 52 个 MIDI 文件，这些文件平均识别率为 94%^[2]。对这 52 个 MIDI 文件两两进行比较，共产生 1326(=52×51/2) 个不同的相似度。利用每个乐曲的旋律进行了相似度计算，求出每个文件的和其它 51 个文件比较的平均相似度。限于篇幅，表 1 列出了平均相似度计算所得较高和较低值的部分结果。所有平均相似度的总平均相似度约为 0.112 342，其中有 3 个文件属于同一首曲子《嘎达梅林》，如果删去其中的 2 个文件，可以得到不同乐曲的总的平均相似度约为 0.108 857。

同一首乐曲的不同数字乐谱图像版本的相似度计算验证，选取了《嘎达梅林》的 3 个不同版本，其相互间的相似度计算结果示于表 2 中。仿真实验结果表明，显然，同一乐曲的不同版本之间的相似度远远高于不同乐曲之间的相似度。这一特性为数字乐谱图像的识别、归类与校正提供了方便，同时也验证了上述算法的有效性和可行性。

表 1 52 首乐曲平均相似度计算的部分结果

乐曲名	乐曲的平均相似度
溜冰园舞曲	0.058 823 529
匈牙利舞曲	0.068 712 702
小星星变奏曲	0.073 262 032
命运交响乐	0.074 425 469
蓝色多瑙河	0.075 980 392
...	...
在月光下	0.164 438 503
小杜鹃	0.166 666 667
法兰西民歌	0.169 340 463
节日的快乐	0.172 715 895
嘎达梅林版本 3	0.200 603 318
嘎达梅林版本 2	0.201 238 39
嘎达梅林版本 1	0.201 608 849
总平均相似度	0.112 342 981

表 2 《嘎达梅林》3 个不同版本之间的相似度

嘎达梅林	版本 1	版本 2	版本 3
版本 1		0.872	0.974
版本 2	0.895		0.921
版本 3	0.974	0.897	

5 结论

大量的优秀音乐作品以不同的形式存在，本文提出了利用 OMR 技术来判断 2 个乐谱图像是不是同一作品的算法，对有效性进行了说明，并通过仿真实验验证了算法的有效性和可行性。为不同数字版本的同一音乐作品的乐谱图像的识别、归类与校正提供了一种有效的实现方法与手段。

参考文献

- Ng K, Barthelemy J, Ong B, et al. Coding Images of Music Sheets[Z]. 2004-02-03. http://www.interactivemusicnetwork.org/documenti/view_document.php?file_id=765.
- Chen G F, Xia S R.. The Study and Prototype System of Printed Music Recognition[C]//Proceedings of 2003 International Conference on Neural Networks and Signal Processing. 2003.
- 卢开橙. 计算机算法导引[M]. 北京: 清华大学出版社, 1996.

(上接第 159 页)

(3) 系统稳定性。当有新成员加入或剔除违规成员 J_p 时，TDC 只需根据 J_p 个人的 $x_p = \text{Red}(J_p)$ 计算 J_p 的私钥为 $\alpha^{f(x_p)} \bmod p'$ ，公钥为 $y_p = g^{\alpha^{f(x_p)}} \bmod p$ ，并在公告牌上公布 (y_p, x_p) 。而群密钥对 $(f(0), y)$ 和每个成员的密钥对 $(\alpha^{f(x_i)} \bmod p', y_i)$ 并未改变。

(4) 多参与性。由于系统每个成员的密钥对始终保持不变，因此任何成员都可以参加群里不同组织，同时为不同的消息签名。

(5) 匿名性和签名的可跟踪性。当发生纠纷时，群秘书可以通过构造的多项式 $h(z)$ 查出签名者的真实身份。

(6) 运算时间分析。由于群秘书 DC 不仅要验证 t 个签名者的部分签名，而且自己也要参与群签名的生成，这无疑增加了他的工作量和签名的执行时间，但它的安全性比原来的方案大大提高了。而且群签名的验证都须用 DC 的公钥，要追查责任先查到他，这就间接要求他必须诚实。因而群秘书在本方案中的地位尽管非常重要，但却不会因此存在不安全隐患。

6 结束语

本文改进后的数字签名方案，由于群秘书的积极参与，

使得方案具有较高的安全性，防止了一些不必要的安全性隐患的发生，具有重要的现实意义。

参考文献

- Desmedt Y, Frankel Y. Shared Generation of Authenticators and Signatures[C]//Proc. of CRYPTO'91. 1991: 457-469.
- Harn L. Group-oriented(t, n)-threshold Digital Signature Scheme and Digital Multi-signature[J]. IEE Proc. of Computers and Digital Techniques, 1994, 141(5): 307-313.
- Li C M, Hwang T, Lee N Y. Threshold Multi-signature Schemes Where Suspected Forgery Implies Traceability of Adversarial Shareholders[C]//Proc. of EUROCRYPT'94, 1995: 194-204.
- 祁明, Harn L. 基于离散对数的若干新型代理签名方案[J]. 电子学报, 2000, 28(11): 114-115.
- Wang C T, Lin C H, Chang C C. Threshold Signature Schemes with Traceable Signers in Group Communications[J]. Computer Comm. 1998, 21(8): 771-776.
- 王晓明, 符方伟. 一种安全的群签名方案[J]. 电子与信息学报, 2003, 42(5): 657-663.

