

[19] 中华人民共和国国家知识产权局



[12] 发明专利申请公布说明书

[21] 申请号 200910307828.0

[51] Int. Cl.

G10H 7/00 (2006.01)

G06K 9/00 (2006.01)

G10L 13/02 (2006.01)

[43] 公开日 2010 年 3 月 3 日

[11] 公开号 CN 101661745A

[22] 申请日 2009.9.28

[21] 申请号 200910307828.0

[71] 申请人 西安理工大学

地址 710048 陕西省西安市金花南路 5 号

[72] 发明人 张海英 郭人通 郝贵青 阳 锋

刘国胜 肖盛杰 张 岩 陈常晋

张 雷 李玉军

[74] 专利代理机构 西安弘理专利事务所

代理人 罗 笛

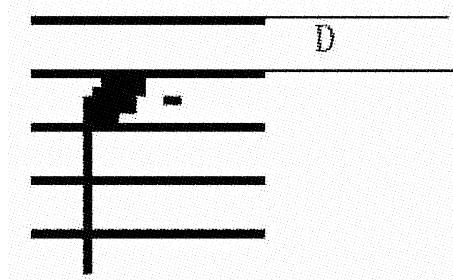
权利要求书 4 页 说明书 10 页 附图 1 页

[54] 发明名称

一种将五线谱图像转换为声音文件的方法

[57] 摘要

本发明公开的一种将五线谱图像转换为声音文件的方法，首先定义音符代码，录制音符的标准声音文件；再采集五线谱图像；然后计算机识别五线谱图像中的小节线、含符干音符、全音符、升降记号、休止符以及谱号，将识别结果转换为音符代码后再存储为声音文件；最后播放声音文件。本发明将五线谱图像转换为声音文件的方法，根据五线谱的知识，对音符的发音规律进行了分类和编码；在图像处理原理的基础上，把人的识别过程转换成对乐谱图像的处理，其识别方法不同于其他图像的识别，识别过程通过 C++ 语言程序设计加以实现。



【权利要求1】一种将五线谱图像转换为声音文件的方法，其特征在于，具体按照以下步骤实施：

步骤1：定义音符代码，录制音符的标准声音文件

给每个音符定义一个代码，代码由四位十进制数组成，第一位为音高，第二位为音名，第三位为音调转换标记，第四位为音长；将每个音符按照规定的音高、音名、音调和音长，录制成标准声音文件，存储在计算机中；

步骤2：采集五线谱图像

用扫描仪或摄像头获取要识别的五线谱图像，以图片格式文件存储在计算机中；

步骤3：计算机识别五线谱图像中的小节线、含符干音符、全音符、升降记号、休止符以及谱号，将识别结果转换为音符代码后再存储为声音文件；

步骤4：播放声音文件

对照步骤1得到的标准声音文件，播放步骤3得到的声音文件，即为要识别的五线谱图像中音符的声音。

【权利要求2】根据权利要求1所述的一种将五线谱图像转换为声音文件的方法，其特征在于，所述的步骤3中将五线谱图像转换为音符代码后再存储为声音文件，具体按照以下步骤实施：

步骤1：图像的二值化

(1) 将彩色图像转换为灰度图像

计算机对RGB三色彩通道位图上的值求图像灰度加权和，采用以下公式：

$$\text{Gray} = \text{Red} * 0.30 + \text{Gre} * 0.59 + \text{Blu} * 0.11$$

得到图像中各个像素的灰度值；

(2) 确定阈值

对灰度图像进行取样，采用大津法对收集到的数据进行处理，获得二值化的阈值，对待处理图像，记t为前景与背景的分割阈值，前景点数占图像比例为w0，平均灰度为u0；背

景点数占图像比例为 w_1 , 平均灰度为 u_1 , 图像的总平均灰度为: $u=w_0*u_0+w_1*u_1$, 从最小灰度值到最大灰度值遍历 t , 当 t 使得值 $g=w_0*w_1*(u_0-u_1)^2$ 最大时 t 即为分割的最佳阈值;

(3) 二值化

灰度值大于或等于阈值的像素, 其灰度值用1表示, 表示属于前景; 否则, 灰度值用0表示, 像素点被排除在物体区域以外, 表示背景或者例外的物体区域;

步骤2: 确定各谱线的位置及相邻线之间的平均间距, 并擦除谱线

(1) 确定谱线位置及相邻线之间的平均间距

对上步得到的二值图进行横向投影, 在投影值列表中找到5个最大的值, 取他们的平均值, 记为 MA , 将阈值设为 $0.5MA$, 投影值大于这个值则判定该条线可能为谱线, 称为候选谱线, 在确定图中各候选谱线后, 对各相邻线的非零线间宽进行统计, 找出占比重最大一组, 将其值确定为单位元长度 D ;

(2) 谱线的擦除

谱线的边界使用 $B2 \times n$ 矩阵进行描述, 其中 n 为谱线的长度, $B1_j$ 为上边界的纵坐标, $B2_j$ 为下边界的纵坐标, 其中 $0 < j \leq n$,

在出现谱线组的位置, 分别对五条谱线所处区域做横向扫描, 当像素点连续为前景的次数大于某设定阈值时, 认为这条线段是构成当前谱线的一部分, 将该线段上各点与 $B2 \times n$ 矩阵中相应的边界坐标比较, 若小于现有上边界或大于现有下边界, 则更新边界信息, 处理完区域内所有的像素点后, 得到完整的横向线边界描述;

步骤3: 图像的分割

若相邻两组谱线的间距大于 $8D$, 则在距边沿谱线 $4D$ 处进行分割; 若相邻两组谱线的间距小于 $8D$, 则在间隔部分的平分线处进行分割;

步骤4: 图像的识别

(1) 小节线及含符干音符的识别

a. 纵向线段的识别

对分割后的图像做纵向投影, 设 S_c 为投影筛选阈值, S_v 为纵向线判定阈值, B_t 为最大断笔次数, D_b 为最大断笔距离, 判定纵向线时, 按以下步骤进行:

若当前列的投影值小于 S_c , 则判定该列上不可能存在纵向线, 这一列直接跳过, 转向下一列进行处理, 若当前列投影值大于 S_c , 转下一步;

从上至下逐点查看权值, 找出存在于当前列的所有离散线段;

若上下相邻两线段的间隔小于 D_b , 那么认为这个间隔是断笔, 可以将这两条线段连为

一条，如此构成的长线段中，存在的断笔次数不能超过Bt；

在线段拼接后，若当前列中最长线段的长度大于Sv，则判定这条线段是一条纵向线；

b. 横梁式符尾的识别

对于相邻的两个矩形块，取它们对应端点坐标的差值，直除后获得连接两个顶端或两个底端的线段，在一定的范围内纵向平移该线段，若线段所处位置上点的权值都为1，则认为在当前位置上存在一条横梁式符尾；

c. 符头的识别

对符杆的左下方和右上方依次测试，若出现了一个类似符头大小的前景区域，则认为此位置存在一个实心符头；在一个符头的范围内做纵向求交测试，若交点次数普遍为2，则认为此位置存在一个空符头；

d. 独立符尾的识别

若符头出现在符杆的左下方，则符尾将出现在符杆的右上方；若符头出现在符杆的右上方，则符尾将出现在符杆的右下方，在符尾可能出现的区域做纵向求交测试，对各列的交点总数进行统计，将出现频率最高的值确定为当前音符的符尾条数；

e. 小节线的识别

在音符的识别之后，查看剩余矩形块的特征，若矩形块的顶端位于第五线，底端位于第一线，且左右两侧0.5D范围内各列的纵向投影值均近似为0，则认为该矩形块为一个小节线；

f. 擦除

识别完毕后，分别擦除符干、小节线、符尾和符头；

(2) 全音符、升降记号、休止符以及谱号的识别

a. 确定外接矩形

对于图中每一个前景的八向连通区域，确定一个与之对应的外接矩形，每一个外接矩形都有4个属性：X轴方向的起始坐标X1、终止坐标X2和Y轴方向的起始坐标Y1、终止坐标Y2，

确定外接矩形的算法如下：

将图像进行纵向投影，对投影值Yi， $0 < i \leqslant$ 图像宽度，进行判断：若Yi>0且Yi-1 = 0，则i为外接矩形的左边界X1；若Yi>0且Yi+1 = 0，则i为外接矩形的右边界X2；

将图像进行横向投影，把X1、X2作为横向投影的两个边界，对投影值Xi， $X1 \leqslant i \leqslant X2$ ，进行判断：若Xi>0且Xi-1=0，则i为外接矩形的上边界Y1；若Xi>0且Xi+1=0，则i为外接矩形

的下边界Y2；

确定了外接矩形后，将这些外接矩形依次与全音符、升降记号、休止符以及谱号的标准外接矩形匹配，进行初步筛选；

b. 细节投影值特征匹配

对外接矩形所确定的区域进行纵向投影，确定一条特征曲线，设定一导数阈值Sd，若曲线上某点处导数绝对值大于等于Sd，则在此处取断点，由各断点可将曲线分为若干段，取两类数值作为细节匹配的特征值：一是曲线段的所在区间的长度，二是区间内曲线所对应函数的平均函数值，由若干个曲线段获得了一组特征序列；

根据上述方法对全音符、升降记号、休止符以及谱号的标准图像进行处理，得到这些音符的标准特征序列，即匹配过程中的参考量；

在处理待匹配对象后，获得其特征序列，待匹配对象的特征序列与标准特征序列的匹配方法如下：

首先求标准特征序列与待测符号特征序列的最长公共子序列；计算匹配度，即最长公共子序列长度除以标准特征序列长度，若匹配度大于设定的阀值，则认为在纵向上该匹配对象与所对应的标准音符匹配；

步骤5：将图像的识别结果转换为音符代码

根据上步得到的带符干音符的符头位置和全音符的位置，确定乐谱中音符的音高和音名；根据上步得到的带符干音符的符尾、符头类型及全音符，确定音符的音长；根据上步得到的休止符的类型，确定休止符的音长；根据上步得到的升降号位置，确定音调转换标记；

步骤6：将音符代码转换为声音文件

将上步识别得到的音符代码，以声音文件的格式存储于计算机中。

一种将五线谱图像转换为声音文件的方法

技术领域

本发明属于图像识别技术领域，涉及一种五线谱识别方法，具体涉及一种将五线谱图像转换为声音文件的方法。

背景技术

目前，国内对于五线谱图像的识别技术还处在初级阶段，通常采用的方法是人工识别五线谱中的音符，然后进行演奏，基于数字图像处理的五线谱图像识别方法还未见涉及，将五线谱图像转换为声音文件并且播放，具有很好的应用前景。

发明内容

本发明的目的是提供一种将五线谱图像转换为声音文件的方法，将图像中的五线谱信息转换为具有某种规律的代码，每一个代码对应一个音符的声音文件，为无声文件转换成有声文件提供基础。

本发明的技术方案是：一种将五线谱图像转换为声音文件的方法，具体按照以下步骤实施：

步骤1：定义音符代码，录制音符的标准声音文件

给每个音符定义一个代码，代码由四位十进制数组成，第一位为音高，第二位为音名，第三位为音调转换标记，第四位为音长；将每个音符按照规定的音高、音名、音调和音长，录制成标准声音文件，存储在计算机中；

步骤2：采集五线谱图像

用扫描仪或摄像头获取要识别的五线谱图像，以图片格式文件存储在计算机中；

步骤3：计算机识别五线谱图像中的小节线、含符干音符、全音符、升降记号、休止符以及谱号，将识别结果转换为音符代码后再存储为声音文件；

步骤4：播放声音文件

对照步骤1得到的标准声音文件，播放步骤3得到的声音文件，即为要识别的五线谱图像中音符的声音。

本发明的特点还在于，步骤3中将五线谱图像转换为音符代码后再存储为声音文件，具体按照以下步骤实施：

步骤1：图像的二值化

(1) 将彩色图像转换为灰度图像

计算机对RGB三色彩通道位图上的值求图像灰度加权和，采用以下公式：

$$\text{Gray} = \text{Red} * 0.30 + \text{Gre} * 0.59 + \text{Blu} * 0.11$$

得到图像中各个像素的灰度值；

(2) 确定阈值

对灰度图像进行取样，采用大津法对收集到的数据进行处理，获得二值化的阈值，对待处理图像，记t为前景与背景的分割阈值，前景点数占图像比例为w0，平均灰度为u0；背景点数占图像比例为w1，平均灰度为u1，图像的总平均灰度为： $u=w0*u0+w1*u1$ ，从最小灰度值到最大灰度值遍历t，当t使得值 $g=w0*w1*(u0-u1)^2$ 最大时t即为分割的最佳阈值；

(3) 二值化

灰度值大于或等于阈值的像素，其灰度值用1表示，表示属于前景；否则，灰度值用0表示，像素点被排除在物体区域以外，表示背景或者例外的物体区域；

步骤2：确定各谱线的位置及相邻线之间的平均间距，并擦除谱线

(1) 确定谱线位置及相邻线之间的平均间距

对上步得到的二值图进行横向投影，在投影值列表中找到5个最大的值，取他们的平均值，记为MA，将阈值设为0.5MA，投影值大于这个值则判定该条线可能为谱线，称为候选谱线，在确定图中各候选谱线后，对各相邻线的非零线间宽进行统计，找出占比重最大一组，将其值确定为单位元长度D；

(2) 谱线的擦除

谱线的边界使用 $B2 \times n$ 矩阵进行描述，其中n为谱线的长度，B1j为上边界的纵坐标，B2j为下边界的纵坐标，其中 $0 < j \leq n$ ，

在出现谱线组的位置，分别对五条谱线所处区域做横向扫描，当像素点连续为前景的次数大于某设定阈值时，认为这条线段是构成当前谱线的一部分，将该线段上各点与 $B2 \times n$ 矩阵中相应的边界坐标比较，若小于现有上边界或大于现有下边界，则更新边界信息，处理完区域内所有的像素点后，得到完整的横向线边界描述；

步骤3：图像的分割

若相邻两组谱线的间距大于8D，则在距边沿谱线4D处进行分割；若相邻两组谱线的间距小于8D，则在间隔部分的平分线处进行分割；

步骤4：图像的识别

(1) 小节线及含符干音符的识别

a. 纵向线段的识别

对分割后的图像做纵向投影，设 S_c 为投影筛选阈值， S_v 为纵向线判定阈值， B_t 为最大断笔次数， D_b 为最大断笔距离，判定纵向线时，按以下步骤进行：

若当前列的投影值小于 S_c ，则判定该列上不可能存在纵向线，这一列直接跳过，转向下一列进行处理，若当前列投影值大于 S_c ，转下一步；

从上至下逐点查看权值，找出存在于当前列的所有离散线段；

若上下相邻两线段的间隔小于 D_b ，那么认为这个间隔是断笔，可以将这两条线段连为一条，如此构成的长线段中，存在的断笔次数不能超过 B_t ；

在线段拼接后，若当前列中最长线段的长度大于 S_v ，则判定这条线段是一条纵向线；

b. 横梁式符尾的识别

对于相邻的两个矩形块，取它们对应端点坐标的差值，直除后获得连接两个顶端或两个底端的线段，在一定的范围内纵向平移该线段，若线段所处位置上点的权值都为1，则认为在当前位置上存在一条横梁式符尾；

c. 符头的识别

对符杆的左下方和右上方依次测试，若出现了一个类似符头大小的前景区域，则认为此位置存在一个实心符头；在一个符头的范围内做纵向求交测试，若交点次数普遍为2，则认为此位置存在一个空符头；

d. 独立符尾的识别

若符头出现在符杆的左下方，则符尾将出现在符杆的右上方；若符头出现在符杆的右上方，则符尾将出现在符杆的右下方，在符尾可能出现的区域做纵向求交测试，对各列的交点总数进行统计，将出现频率最高的值确定为当前音符的符尾条数；

e. 小节线的识别

在音符的识别之后，查看剩余矩形块的特征，若矩形块的顶端位于第五线，底端位于第一线，且左右两侧0.5D范围内各列的纵向投影值均近似为0，则认为该矩形块为一个小节线；

f. 擦除

识别完毕后，分别擦除符干、小节线、符尾和符头；

(2) 全音符、升降记号、休止符以及谱号的识别

a. 确定外接矩形

对于图中每一个前景的八向连通区域，确定一个与之对应的外接矩形，每一个外接矩形

都有4个属性：X轴方向的起始坐标X1、终止坐标X2和Y轴方向的起始坐标Y1、终止坐标Y2，

确定外接矩形的算法如下：

将图像进行纵向投影，对投影值 Y_i , $0 < i \leq$ 图像宽度，进行判断：若 $Y_i > 0$ 且 $Y_{i-1} = 0$ ，则 i 为外接矩形的左边界X1；若 $Y_i > 0$ 且 $Y_{i+1} = 0$ ，则 i 为外接矩形的右边界X2；

将图像进行横向投影，把X1、X2作为横向投影的两个边界，对投影值 X_i , $X_1 \leq i \leq X_2$ ，进行判断：若 $X_i > 0$ 且 $X_{i-1}=0$ ，则 i 为外接矩形的上边界Y1；若 $X_i > 0$ 且 $X_{i+1}=0$ ，则 i 为外接矩形的下边界Y2；

确定了外接矩形后，将这些外接矩形依次与全音符、升降记号、休止符以及谱号的标准外接矩形匹配，进行初步筛选；

b. 细节投影值特征匹配

对外接矩形所确定的区域进行纵向投影，确定一条特征曲线，设定一导数阈值Sd，若曲线上某点处导数绝对值大于等于Sd，则在此处取断点，由各断点可将曲线分为若干段，取两类数值作为细节匹配的特征值：一是曲线段的所在区间的长度，二是区间内曲线所对应函数的平均函数值，由若干个曲线段获得了一组特征序列；

根据上述方法对全音符、升降记号、休止符以及谱号的标准图像进行处理，得到这些音符的标准特征序列，即匹配过程中的参考量；

在处理待匹配对象后，获得其特征序列，待匹配对象的特征序列与标准特征序列的匹配方法如下：

首先求标准特征序列与待测符号特征序列的最长公共子序列；计算匹配度，即最长公共子序列长度除以标准特征序列长度，若匹配度大于设定的阀值，则认为在纵向上该匹配对象与所对应的标准音符匹配；

步骤5：将图像的识别结果转换为音符代码

根据上步得到的带符干音符的符头位置和全音符的位置，确定乐谱中音符的音高和音名；根据上步得到的带符干音符的符尾、符头类型及全音符，确定音符的音长；根据上步得到的休止符的类型，确定休止符的音长；根据上步得到的升降号位置，确定音调转换标记；

步骤6：将音符代码转换为声音文件

将上步识别得到的音符代码，以声音文件的格式存储于计算机中。

本发明将五线谱图像转换为声音文件的方法，根据五线谱的知识，对音符的发音规律进行了分类和编码；在图像处理原理的基础上，把人的识别过程转换成对乐谱图像的处理，其识别方法不同于其他图像的识别，识别过程通过C++语言程序设计加以实现。

附图说明

图1为本发明方法中相邻线之间的平均间距示意图。

具体实施方式

下面结合附图和具体实施方式对本发明进行详细说明。

本发明方法对应的识别对象为谱线、小节线、谱号、音符、休止符、升降记号。谱线为第一组识别对象，由五个一组的等间距横向长线组成，特征最为明显，且贯穿所有其他识别对象，因此应最先得到处理；小节线和具有符干的音符为第二组识别对象，小节线是一条或两条一定长度的纵向线，具有符干的音符，其符干可认为是一条不定长的纵向线，故可根据这些纵向线将这一组对象从图像中分离出来，因这一组对象一般在乐谱中出现极为频繁，将其优先识别可在很大程度上减少后续识别中的计算量及干扰；全音符、升降记号、休止符以及谱号为第三组识别对象，他们的共同特征是形态固定，但形状较为复杂，识别中易受干扰，故将这一组放到最后处理。

本发明将五线谱图像转换为声音文件的方法，具体按照以下步骤实施：

步骤1：定义音符代码，录制音符的标准声音文件

给每个音符定义一个代码，代码由四位十进制数组成，第一位为音高，第二位为音名，第三位为音调转换标记，第四位为音长；将每个音符按照规定的音高、音名、音调和音长，录制成标准声音文件，以“音符代码.wav”格式存储在计算机中；

步骤2：采集五线谱图像

用扫描仪或摄像头获取要识别的五线谱图像，以图片格式文件存放在计算机中；

步骤3：将五线谱图像转换为声音文件

1. 图像的二值化

(1) 将彩色图像转换为灰度图像

计算机对RGB三色彩通道位图上的值求图像灰度加权和，采用经典转换公式：

$$\text{Gray} = \text{Red} * 0.30 + \text{Gre} * 0.59 + \text{Blu} * 0.11 \quad (1)$$

得到图像中各个像素的灰度值；

(2) 确定阈值

对灰度图像进行取样，采用大津法对收集到的数据进行处理，获得二值化的阈值。对待处理图像，记t为前景与背景的分割阈值，前景点数占图像比例为w0，平均灰度为u0；背景点数占图像比例为w1，平均灰度为u1。图像的总平均灰度为： $u=w0*u0+w1*u1$ 。从最小灰度值到最大灰度值遍历t，当t使得值 $g=w0*(u0-u)^2+w1*(u1-u)^2$ 最大时t即为分割的最佳阈

值。直接应用大津法计算量较大，这里采用其等价公式 $g=w0*w1*(u0-u1)^2$ 。

二值化

灰度值大于或等于阈值的像素，其灰度值用1表示，表示属于前景；否则，灰度值用0表示，像素点被排除在物体区域以外，表示背景或者例外的物体区域，由此生成二值图。

2. 确定各谱线的位置及相邻线之间的平均间距，并擦除谱线

(1) 确定谱线位置及相邻线之间的平均间距

对上步得到的二值图进行横向投影，在投影值列表中找到5个最大的值，取他们的平均值，记为MA。将阈值设为0.5MA（通过对多幅图片处理结果的分析，将阈值定为此值效果较为理想），投影值大于这个值则判定该条线可能为谱线，这里称为候选谱线。在确定图中各候选谱线后，对各相邻线的非零线间宽进行统计，找出占比重最大一组，将其值确定为单位元长度D，如图1所示。

为排除图像中的干扰线，五线组需要满足以下附加条件：

- a. 若当前候选谱线为整张图像的第一条线，或与上一条相邻线的间距不为D，或继一个五线组之后出现，则认为此线为当前五线组的第一条线；
- b. 若当前候选谱线不是该组的第一条线且它到上一条相邻线的间距为D，则认为当前线是五线组中的一员；
- c. 若一组线的成员数目达到5，则可判定该组线为一个五线组。

在实际乐谱中，各组谱线往往存在不等长的情况，第一行或最后一行的谱线组有可能相对较短。这些谱线组若短于判定阈值，则会被直接丢弃，这将造成严重的错误，故需将以上算法做如下改进：对被丢弃图像块的大小进行检查，若认为该块图像有可能含有被丢弃的谱线组，则对于该块图像重复调用上述识别方法。

(2) 谱线的擦除

通过观察二值图，发现构成谱线组的谱线不是笔直的，而是由多段较短的横向线相互交错拼接组成。故在去除五线时需要根据这些短线段所处位置不同而做出不同的反应。具体方法如下：

谱线的边界使用 $B2 \times n$ 矩阵进行描述，其中n为谱线的长度， $B1_j$ 为上边界的纵坐标， $B2_j$ 为下边界的纵坐标，其中 $0 < j \leq n$ 。

在出现谱线组的位置，分别对五条谱线所处区域做横向扫描。当像素点连续为前景的次数大于某设定阈值时，认为这条线段是构成当前谱线的一部分。将该线段上各点与 $B2 \times n$ 矩阵中相应的边界坐标比较，若小于现有上边界或大于现有下边界，则更新边界信息。处理完

区域内所有的像素点后，将得到完整的横向线边界描述。

直接根据五线的边界信息来消除五线会打断音符，因此在擦除五线前，需要判断谱线上各点是否是其他音符的组成部分，若是，应予以保留。判断处于($j, B1j - 1$)位置或($j, B2j + 1$)位置上像素点是否前景点，若是，则($j, B1j$)到($j, B2j$)的线段予以保留。

3. 图像的分割

为防止在识别过程中各行乐谱相互干扰，需对二值图进行分割。分割方法如下：若相邻两组谱线的间距大于 $8D$ ，则在距边沿谱线 $4D$ 处进行分割；若相邻两组谱线的间距小于 $8D$ ，则在间隔部分的平分线处进行分割。

4. 图像的识别

(1) 小节线及符干音符的识别

a. 纵向线段的识别

所需要的纵向线是一些具有一定长度和连续性的线段。对分割后的图像做纵向投影，设 S_c 为投影筛选阈值， S_v 为纵向线判定阈值， B_t 为最大断笔次数， D_b 为最大断笔距离。判定纵向线时，按以下步骤进行：

若当前列的投影值小于 S_c ，则判定该列上不可能存在纵向线，这一列直接跳过，转向下一列进行处理。若当前列投影值大于 S_c ，转下一步；

从上至下逐点查看权值，找出存在于当前列的所有离散线段；

若上下相邻两线段的间隔小于 D_b ，那么认为这个间隔是断笔，可以将这两条线段连为一条，如此构成的长线段中，存在的断笔次数不能超过 B_t ；

在线段拼接后，若当前列中最长线段的长度大于 S_v ，则判定这条线段是一条纵向线。

至此，可以得到一些零散的纵向线，这些线大体可分为三类：第一类是构成符杆或小节线的元素；第二类是依附于符杆两侧的，由符头和符尾构成的纵向线；第三类是满足判定条件的干扰图像。其中的第一类是我们所需要的。

第一类纵向线的特征是：在每一组中，它们彼此相邻，且长度和起始位置相似。可以根据这个特征将各组纵向线分别连接起来，构成一个个矩形面，这些矩形面即是小节线或符干。

第二类纵向线虽然和第一类纵向线是相邻的，但由于长度不同，在投影图中，第二类线与第一类线之间的过度极不平缓，在过度处图像的导数急剧变化，故可据此将第二类纵线从第一类纵线中剥离开。

第三类纵向线接受第一类线相同的处理。由第三类线所拼接成的矩形中，一部分的宽度

或长度与符杆及小节线不符，因此这些线可以被分离出来。若宽度及长度都相符，则予以保留，在以后的步骤中处理。

b. “横梁式”符尾的识别

对于独立的带有符尾的音符，若符头出现在符杆的左下方，则符尾将出现在符杆的右上方；若符头出现在符杆的右上方，则符尾将出现在符杆的右下方。但“横梁式”符尾并不满足上述规律，它们只是处于符头的相反端，连接相邻的两个符干。同时，在图像并不十分清晰的情况下，短的“横梁”与符头十分相似，这将给符头的识别带来障碍。出于对以上两个原因的考虑，“横梁式”符尾应当先于符头及独立符尾得到处理。

对于相邻的两个矩形块，取它们对应端点坐标的差值，直除后可获得连接两个顶端或两个底端的线段。在一定的范围内纵向平移该线段，若线段所处位置上点的权值都为1，则可认为在当前位置上存在一条“横梁式”符尾。

c. 符头的识别

符头有实心符头和空心符头两种，出现位置为符杆的左下方或右上方。对符杆的左下方和右上方依次测试，若出现了一个类似符头大小的前景区域，则认为此位置存在一个实心符头；在一个符头的范围内做纵向求交测试，若交点（0-1?0视为一个交点）次数普遍为2，则认为此位置存在一个空符头。虽然16分音符（含两条符尾）的符尾部分也满足空符头的判定条件，但并不会产生误判。这是因为实心符头先于空符头进行识别，而16分音符的符头恰是实心的。

d. 独立符尾的识别

在确定符头之后，即可知道符尾的位置。若符头出现在符杆的左下方，则符尾将出现在符杆的右上方；若符头出现在符杆的右上方，则符尾将出现在符杆的右下方。在符尾可能出现的区域做纵向求交测试，对各列的交点总数进行统计，将出现频率最高的值确定为当前音符的符尾条数。

e. 小节线的识别

在音符的识别之后，查看剩余矩形块的特征。若矩形块的顶端位于第五线，底端位于第一线，且左右两侧0.5D范围内各列的纵向投影值均近似为0，则认为该矩形块为一个小节线。

f. 擦除

上述识别完毕后，分别擦除符干、小节线、符尾和符头。

（2）全音符、升降记号、休止符以及谱号的识别

a. 确定外接矩形

为了能够方便准确的对上述音符进行识别，这里采用两级特征匹配的方法。首先根据外接矩形的尺寸特征进行初步筛选。对于图中每一个前景的八向连通区域，都可以确定一个与之对应的外接矩形。每一个外接矩形都有4个属性：X轴方向的起始坐标X1、终止坐标X2和Y轴方向的起始坐标Y1、终止坐标Y2。

确定外接矩形的算法如下：

将图像进行纵向投影，对投影值 Y_i ($0 < i \leq \text{图像宽度}$) 进行判断：若 $Y_i > 0$ 且 $Y_{i-1} = 0$ ，则 i 为外接矩形的左边界X1；若 $Y_i > 0$ 且 $Y_{i+1} = 0$ ，则 i 为外接矩形的右边界X2。

将图像进行横向投影，把X1、X2作为横向投影的两个边界，对投影值 X_i ($X_1 \leq i \leq X_2$) 进行判断：若 $X_i > 0$ 且 $X_{i-1} = 0$ ，则 i 为外接矩形的上边界Y1；若 $X_i > 0$ 且 $X_{i+1} = 0$ ，则 i 为外接矩形的下边界Y2。

由于存在断笔，一个整体音符有可能被分成多个部分，而每个部分都确定了一个外接矩形。为了解决这个问题，需要把本属于一个整体的多个外接矩形合并。设最大断笔距离为d，相邻两个外接矩形为R1、R2。若R1与R2的相邻边界间距离小于d，则将它们合并起来。

确定了外接矩形后，将这些外接矩形依次与全音符、升降记号、休止符以及谱号的标准外接矩形匹配，进行初步筛选。

b. 细节投影值特征匹配

对外接矩形所确定的区域进行纵向投影，这些投影值可以确定一条特征曲线，设定一导数阈值 S_d ，若曲线上某点处导数绝对值大于等于 S_d ，则在此处取断点，由各断点可将曲线分为若干段。这里取两类数值作为细节匹配的特征值：一是曲线段的所在区间的长度，二是区间内曲线所对应函数的平均函数值。这样，就由若干个曲线段获得了一组特征序列。

根据上述方法对全音符、升降记号、休止符以及谱号的标准图像进行处理，可得到这些音符的标准特征序列，即匹配过程中的参考量。

在处理待匹配对象后，可获得其特征序列。待匹配对象的特征序列与标准特征序列的匹配方法如下：

首先求标准特征序列与待测符号特征序列的最长公共子序列；计算匹配度，即最长公共子序列长度除以标准特征序列长度。若匹配度大于设定的阀值，则认为在纵向上该匹配对象与所对应的标准音符匹配。

如果对于匹配度有更高的要求，还可以在纵向投影的匹配后，进行横向投影的细节匹配。

5. 将图像的识别结果转换为音符代码

根据上步得到的带符干音符的符头位置和全音符的位置，确定乐谱中音符的音高和音名；根据上步得到的带符干音符的符尾、符头类型及全音符，确定音符的音长；根据上步得到的休止符的类型，确定休止符的音长；根据上步得到的升降号位置，确定音调转换标记。

6. 将音符代码转换为声音文件

将上步识别得到的音符代码，加上“.wav”形成一个字符串，即声音文件名，以声音文件的格式存储于计算机中。

步骤4：播放声音文件

对照步骤1得到的标准声音文件，播放步骤3得到的声音文件，即为要识别的五线谱图像中音符的声音。

本发明将五线谱图像转换为声音文件的方法，根据五线谱的知识，对音符的发音规律进行了分类和编码；在图像处理原理的基础上，把人的识别过程转换成对乐谱图像的处理，其识别方法不同于其他图像的识别，识别过程通过C++语言程序设计加以实现。

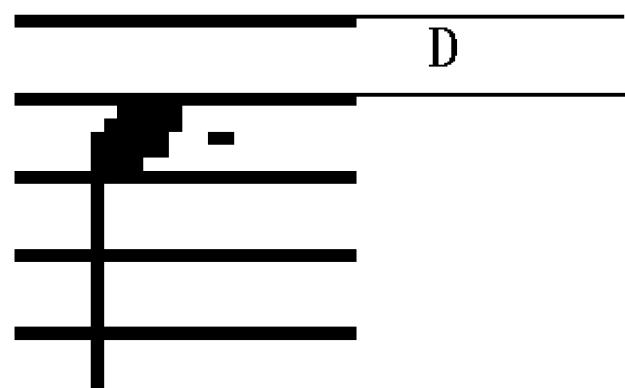


图1