

纸质简谱图像的分割、倾斜校正及 音符歌词提取

邓翔宇, 杨雅涵

(西北师范大学物理与电子工程学院, 甘肃兰州 730070)

摘要: 乐谱图像的自动分割、倾斜校正是乐谱识别过程中的关键技术, 各种计算机光学乐谱识别技术在乐谱图像的数字化中有着广泛的应用, 但对于乐谱中简谱的识别一直鲜有研究. 本文针对人工拍摄条件下光照不理想的简谱图像, 提出一种基于 PCNN(脉冲耦合神经网络) 和 DNN(深度神经网络) 相结合的分块简谱图像自动分割算法, 该方法根据简谱图像灰度分布特征对图像进行自适应分块处理, 依据每个分块的灰度特征与 PCNN 最佳迭代次数之间的关系构造合适的 DNN 神经网络, 从而实现了最优分割图的自适应选取; 进一步利用最优分割图像中音符小节线的水平投影, 提出一种双尺度下降法实现了简谱图像的倾斜校正; 提出去边垂直投影法和连通域距离判断法实现了简谱图像中音符及歌词的提取. 实验仿真结果表明: 本文算法对复杂光照条件下的简谱图像处理都具有较好的鲁棒性, 同时表现出更高的效率.

关键词: PCNN 分块分割; DNN 最优选择; 双尺度下降; 倾斜校正; 去边垂直投影; 音词符提取

中图分类号: TP183 **文献标识码:** A **文章编号:** 0372-2112 (2021)04-0716-13

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.12263/DZXB.20200209

Segmentation, Tilt Correction and Note Lyrics Extraction of Paper Numbered Musical Notation Images

DENG Xiang-yu, YANG Ya-han

(College of Physics and Electronic Engineering, Northwest Normal University, Lanzhou, Gansu 730070, China)

Abstract: Automatic segmentation and tilt correction of music images are key techniques in music recognition. Various computer optical music score recognition technologies have been widely used in the digitization of music score images, but there has been little research on the recognition of numbered musical notation. In this paper, an automatic image segmentation algorithm based on PCNN(Pulse Coupled Neural Networks) and DNN(Deep Neural Networks) is proposed to solve the problem of a variety of lighting conditions. The image is processed by adaptive block processing according to the gray scale distribution of the spectral image and analyze the relationship between the gray scale characteristics of each block and the optimal PCNN iteration time, construct an appropriate DNN neural network to realize the adaptive selection of optimal segmentation graph. Further using the horizontal projection of bar lines, we propose a dual-scale descent method to realize the skew correction of numbered musical notation image. We propose an edgeless vertical projection method and connected domain distance judgment method to extract the note and lyrics from numbered musical notation image. Simulation experiments show that the proposed algorithm exhibits better robustness for numbered musical notation image under complex illumination conditions and faster speed.

Key words: PCNN block segmentation; DNN optimal selection; double scale descent; tilt correction; removable edges vertical projection; extraction of notes and lyrics

1 引言

音乐是人类的第二语言,是人类精神生活中一个重要且最具魅力的领域^[1],其不仅仅对传统文化具有保护和传承的重要意义,同时也在应用领域也具有潜在的市场推广价值.而乐谱又是传承、推广音乐的一种重要载体.国际上乐谱的种类分为五线谱与简谱,其中简谱本身具有简单易学和排印方便的特点,因此在世界上得到了大力的推广与发展.光学乐谱识别技术主要有:基于深度学习的识别方法^[2,3],基于人工神经网络的识别方法^[4]以及基于乐谱特征的识别方法^[5-10].如:Calvozaragoza 等利用卷积神经网络对手写体五线谱符号进行识别研究^[2],同时基于 CRNN 神经网络,提出一种端到端的方法实现了五线谱图像的乐谱识别^[3];Calvozaragoza 等利用隐马尔克夫模型和神经网络相结合的方法对手写五线谱进行识别,但效果受限于对乐谱音符的预分割处理^[4].基于乐谱特征的识别也得到了国内外学者的广泛研究,主要包括乐谱图像的分割、倾斜校正、音符歌词提取三个预处理环节以及识别环节.其中乐谱图像的预处理是基于乐谱特征识别方法的核心环节,也是决定后续音符及歌词识别准确率的重要因素.如:陈根方等将数学形态学应用于五线谱乐谱的提取获得了较好的效果,但处理的图像均为标准下载原图,缺乏对真实拍摄乐谱图处理的详尽分析^[5].乐谱倾斜校正的方法主要有 Hough 变换法^[6,7]、Rado 变换法^[8]等.如:Jiang 等利用图像中直线水平线、垂直线的平均斜率值实现了对标准简谱图的倾斜校正^[9];高强利用霍夫变换检测简谱图像中倾角,再用旋转矩阵对倾角进行校正,并在全局阈值分割法的基础上提出一种简谱乐理符号识别的方法^[10];Rebelo 等根据乐谱图像中的横线特征,提出直线平衡优化算法实现了五线谱图像的倾斜校正^[11].现有的倾斜校正算法大多适用于五线谱图像中,在简谱图像的倾斜校正中存在一定的局限性.针对于音符歌词的提取,现有研究主要是基于音符的结构特征来实现的,如:Luth 首先对五线谱中的符号进行边缘检测,然后将乐谱分成多层次结构再利用符号图像特征实现了五线谱信息的提取^[12];刘晓翔提出基于图段拓扑的谱线删除方法和双向游程编码相结合使用的符号分割方法,解决了音符适应性差、分割结果不完整等缺陷,但是其主要研究依旧集中在五线谱上^[13].

总之,现有的研究大都集中在五线谱图像的数字应用,也有一些研究侧重于标准简谱图的识别研究,但都具有各自的局限性,尤其是对实际人工拍摄的纸质简谱数字化鲜有研究.针对于简谱图像识别过程中的预处理环节,并基于我们前期对 PCNN(Pulse Coupled

Neural Network)应用于图像分割的研究结果^[14-16],本文分别就简谱图像的分割、倾斜校正、音符字符提取进行了研究.首先,设计了一种简谱图像自动分块的策略,并利用 PCNN 实现了对简谱图像的最优分割,进一步将每个分块图像的灰度统计特征作为训练样本,将每个分块的 PCNN 最佳分割迭代次数作为标签,实现了基于 DNN 的最优简谱分割图像的自适应选取;其次,根据最优分割图中小节线的水平投影值,提出了一种双尺度下降法实现了简谱图像的倾斜校正;最后,利用连通域滤除和去边垂直投影的方法,对简谱中的音符和歌词进行了定位和提取.

2 简谱图的自动分割和最优选取

2.1 基于 PCNN 的简谱图像分块分割

PCNN: (脉冲耦合神经网络)是一种单层的、无需训练的第三代人工神经网络,在图像处理领域有着广泛的应用.1990年,Eckhorn 等人提出基于猫的视觉皮层神经元信号传导特性的神经网络模型^[17];1999年,Johnson 等人将其改进成适合于图像处理的模型,并命名为 PCNN^[18].由于其对图像中邻近区域的相近灰度值像素具有同步发放的特性,利用这一特性便可容易地将图像中的亮、暗区域分离出来^[19].而对于纸质简谱图像,无论人为拍摄条件如何,音符及歌词部分相比底色背景都具有明显的亮或者暗特征,因此,PCNN 网络是适合于简谱图像分割的.PCNN 网络由多个基本神经元模型组合而成,其神经元模型如图 1 所示.

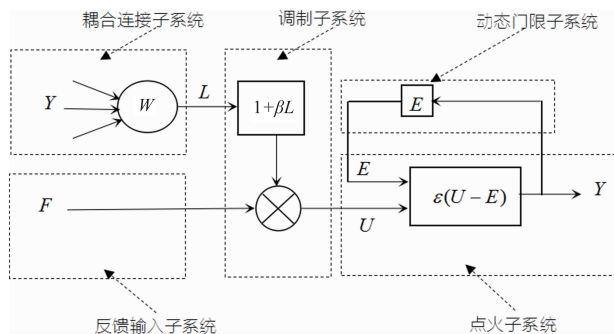


图1 脉冲耦合神经元模型

图 1 所示的神经元模型中,各子系统均可以描述成一个离散系统模型^[20-23],在用于数字图像处理时往往可以简化为式(1)~(5)所示:

$$F_{ij}(n) = S_{ij} \quad (1)$$

$$L_{ij}(n) = V_L \sum W_{ijkl} Y_{ijkl}(n-1) \quad (2)$$

$$U_{ij}(n) = F_{ij}(n) (1 + \beta L_{ij}(n)) \quad (3)$$

$$E_{ij}(n) = e^{-\alpha} E_{ij}(n-1) + V_E Y_{ij}(n-1) \quad (4)$$

$$Y_{ij}(n) = \varepsilon [U_{ij}(n) - E_{ij}(n)] \quad (5)$$

其中,式(1)为反馈输入子系统,式(2)为耦合连接子系

统,式(3)为调制子系统,式(4)为动态门限子系统,式(5)为点火子系统.下标 i, j 表示数字图像中的某一个像素位置; k, l 表示中心像素的邻域像素; \mathbf{W} 表示中心像素和邻域像素的连接权矩阵; V_L, V_E, α_E 分别表示对应子系统的连接权放大系数和迭代衰减时间常数. β 为调制子系统的内部活动项连接系数, S_{ij} 为神经元对应像素点的灰度值.

图2是PCNN采用我们在文献[24]中提出的参数自

适应设置下,简谱标准下载图的PCNN迭代输出取反图,其中,第9次迭代对目标提取更准确,细节更清晰视为最优分割图.图3是人工拍摄简谱图的PCNN迭代过程.可见,简谱图像下方受光照影响整体偏暗,从图像整体分割情况来看,虽然第10次迭代输出一定程度上完成了分割处理,但图像下方出现了过分割的情况.因此,对于复杂光照条件下人工拍摄的简谱图像,直接利用PCNN对其进行迭代输出是无法获得全局最优分割图的.

1=F 4/4
草原之恋
依恋 柔情地 (李庆芳演唱) 刘云海词 韩洪森/房天菊曲

天 蓝 蓝 鸿雁 飞翔 彩云 间 草 青 青
心 依 恋 陶醉 你的 山水 间 茶 一 杯

花 正 艳 骏马 奔驰在草 原 你的 美 丽
酒 一 碗 牧歌 回荡在天 边 你的 魅 力

迷 迷 那 双 眼 让我 流 连 流 连 恋 恋 你的 柔 情
让 我 迷 恋 爱 你 心 甘 心 甘 情 愿 你的 柔 情

拨 动 心 弦 在 你 的 爱 河 里 爱 河 里 淌 流
净 朴 自 然 在 你 的 怀 抱 里 怀 抱 里 入 眠

怀 抱 里 入 眠 (古号记谱)

原图

1=F 4/4
草原之恋
依恋 柔情地 (李庆芳演唱) 刘云海词 韩洪森/房天菊曲

天 蓝 蓝 鸿雁 飞翔 彩云 间 草 青 青
心 依 恋 陶醉 你的 山水 间 茶 一 杯

花 正 艳 骏马 奔驰在草 原 你的 美 丽
酒 一 碗 牧歌 回荡在天 边 你的 魅 力

迷 迷 那 双 眼 让我 流 连 流 连 恋 恋 你的 柔 情
让 我 迷 恋 爱 你 心 甘 心 甘 情 愿 你的 柔 情

拨 动 心 弦 在 你 的 爱 河 里 爱 河 里 淌 流
净 朴 自 然 在 你 的 怀 抱 里 怀 抱 里 入 眠

怀 抱 里 入 眠 (古号记谱)

第8次迭代 熵为0.31235

1=F 4/4
草原之恋
依恋 柔情地 (李庆芳演唱) 刘云海词 韩洪森/房天菊曲

天 蓝 蓝 鸿雁 飞翔 彩云 间 草 青 青
心 依 恋 陶醉 你的 山水 间 茶 一 杯

花 正 艳 骏马 奔驰在草 原 你的 美 丽
酒 一 碗 牧歌 回荡在天 边 你的 魅 力

迷 迷 那 双 眼 让我 流 连 流 连 恋 恋 你的 柔 情
让 我 迷 恋 爱 你 心 甘 心 甘 情 愿 你的 柔 情

拨 动 心 弦 在 你 的 爱 河 里 爱 河 里 淌 流
净 朴 自 然 在 你 的 怀 抱 里 怀 抱 里 入 眠

怀 抱 里 入 眠 (古号记谱)

第10次迭代熵为0.31657

1=F 4/4
草原之恋
依恋 柔情地 (李庆芳演唱) 刘云海词 韩洪森/房天菊曲

天 蓝 蓝 鸿雁 飞翔 彩云 间 草 青 青
心 依 恋 陶醉 你的 山水 间 茶 一 杯

花 正 艳 骏马 奔驰在草 原 你的 美 丽
酒 一 碗 牧歌 回荡在天 边 你的 魅 力

迷 迷 那 双 眼 让我 流 连 流 连 恋 恋 你的 柔 情
让 我 迷 恋 爱 你 心 甘 心 甘 情 愿 你的 柔 情

拨 动 心 弦 在 你 的 爱 河 里 爱 河 里 淌 流
净 朴 自 然 在 你 的 怀 抱 里 怀 抱 里 入 眠

怀 抱 里 入 眠 (古号记谱)

第7次迭代 熵为0.29016

1=F 4/4
草原之恋
依恋 柔情地 (李庆芳演唱) 刘云海词 韩洪森/房天菊曲

天 蓝 蓝 鸿雁 飞翔 彩云 间 草 青 青
心 依 恋 陶醉 你的 山水 间 茶 一 杯

花 正 艳 骏马 奔驰在草 原 你的 美 丽
酒 一 碗 牧歌 回荡在天 边 你的 魅 力

迷 迷 那 双 眼 让我 流 连 流 连 恋 恋 你的 柔 情
让 我 迷 恋 爱 你 心 甘 心 甘 情 愿 你的 柔 情

拨 动 心 弦 在 你 的 爱 河 里 爱 河 里 淌 流
净 朴 自 然 在 你 的 怀 抱 里 怀 抱 里 入 眠

怀 抱 里 入 眠 (古号记谱)

第9次迭代 熵为0.31495

1=F 4/4
草原之恋
依恋 柔情地 (李庆芳演唱) 刘云海词 韩洪森/房天菊曲

天 蓝 蓝 鸿雁 飞翔 彩云 间 草 青 青
心 依 恋 陶醉 你的 山水 间 茶 一 杯

花 正 艳 骏马 奔驰在草 原 你的 美 丽
酒 一 碗 牧歌 回荡在天 边 你的 魅 力

迷 迷 那 双 眼 让我 流 连 流 连 恋 恋 你的 柔 情
让 我 迷 恋 爱 你 心 甘 心 甘 情 愿 你的 柔 情

拨 动 心 弦 在 你 的 爱 河 里 爱 河 里 淌 流
净 朴 自 然 在 你 的 怀 抱 里 怀 抱 里 入 眠

怀 抱 里 入 眠 (古号记谱)

第11次迭代熵为0.33496

图2 标准简谱下载图PCNN迭代输出过程



图3 人工拍摄简谱图PCNN迭代输出过程

根据 PCNN 的同步发放特性可知,简谱图像中音符及歌词的像素灰度值决定了音符及歌词部分像素被熄灭的迭代时刻.由人工拍摄简谱图像的原图可知,简谱图下方受光照影响偏暗,其灰度值偏小,其值越小音符及歌词被完全熄灭的时刻就越早,简谱区域最小值到最大值的过渡越平缓即对比度越小,背景像素对音符像素的耦合作用就越弱,在迭代过程中同步熄灭的趋势就越缓慢.为了保证偏暗部分熄灭时刻不影响其他部分熄灭时刻,只有将偏暗部分的图像与整体图像分开单独处理,才能得到简谱图的全局最优分割.因此,对于复杂光照条件下的简谱图,本文提出对简谱图分块之后再行 PCNN 迭代的处理方式,用来避免整幅简谱图直接做 PCNN 处理带来的局部欠分割与过分割问题.图像分块大小的算法可表示为式(6)~(9):

$$Image_1 = a \tag{6}$$

$$Image_b = b \tag{7}$$

$$Image1_1 = k_i^{\max-l} - k_i^{\min-l} \tag{8}$$

$$Image1_b = k_i^{\max-b} - k_i^{\min-b} \tag{9}$$

其中, a, b 分别表示原图 $Image$ 的长和宽. 根据前期的研究发现,经过 PCNN 多次迭代便可得到图像的过分割图,即图像原本偏暗的部分全部熄灭,由此可清晰的判别出熄灭部分是过分割图像中最大的连通域 i ,即连通域 i 是图像中偏暗的部分. $k_i^{\max-l}$ 表示第 i 个连通域的最大行, $k_i^{\min-l}$ 表示第 i 个连通域的最小行. $k_i^{\max-b}$ 表示第 i 个连通域的最大列, $k_i^{\min-b}$ 表示第 i 个连通域的最小列. $Image1_1$ 则表示连通域 i 的长, $Image1_b$ 表示连通域 i 的宽. 若 $Image1_1 < \frac{a}{3}$, 将原图偏暗部分图像分块大小为 $Image1_1 \times \frac{b}{3}$, 若 $\frac{a}{3} \leq Image1_1 < \frac{2a}{3}$, 原图偏暗部分图像分块大小为 $\frac{Image1_1}{2} \times \frac{b}{3}$, 若 $\frac{2a}{3} \leq Image1_1 < a$, 原图偏暗部分图像分块大小为 $\frac{Image1_1}{3} \times \frac{b}{3}$, 若原图剩余部分大小不满足上述分块大小,则分成 $(a - Image1_1) \times \frac{b}{3}$. 这样定义分块的大小是避免图像中光照偏暗部分与光照正

常部分相冗余,解决了分块图像出现过分割或欠分割的问题,因此可实现简谱图像的全局最优分割. 图像自适应分块过程如图 4 所示.

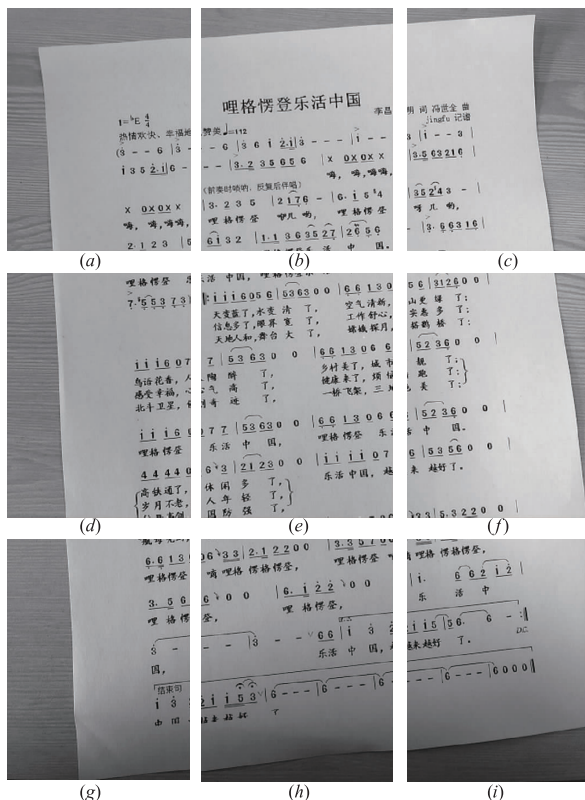


图4 人工拍摄简谱图分块处理

图 4(a) ~ (i) 是人工拍摄简谱图的自适应分块结果,共分为 9 块,其中图 4(a) ~ (f) 拍摄清晰,文字信息较背景底色具有亮度高对比度大的特点,视觉上最为清晰;图 4(g) ~ (i) 是简谱原图中下方偏暗的部分,但对每个分块而言,整个背景底色灰度基本接近,从而可以确保音符和歌词部分在 PCNN 迭代过程中的同步发放. 图 5(a) ~ (i) 则为每一分块简谱图的最优分割图,可见,得到的分割图像目标清晰,没有出现过分割或欠分割情况. 其中图 5(g) ~ (i) 即使原图图像整体偏暗,但基于 PCNN 的脉冲发放特性分析可知,分块图像中像素本身的灰度值决定了其在网络迭代过程中的灭火时刻,灰度值越小灭火时刻就越早,而邻域像素则通过耦合连接作用对像素本身的灭火时刻进行微调,从而使得邻域内灰度值相近的像素同步灭火^[22],因此便也得到了最优分割图,解决了整幅简谱图直接做 PCNN 处理带来的欠分割与过分割问题.

图 6(a) 为真实光照条件下人工拍摄的简谱原图,整体对比度分布不均匀,图像底部偏暗;图 6(b) 为整幅图像利用 PCNN 算法得到的最优分割,出现了图像底部过分割的情况;图 6(c) 为本文提出的分块图像 PCNN

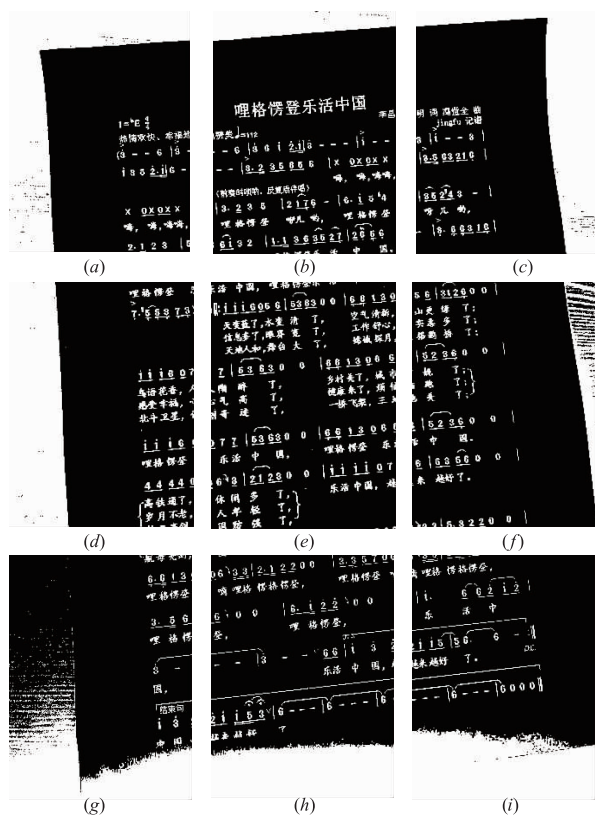


图5 分块图像PCNN最优分割图

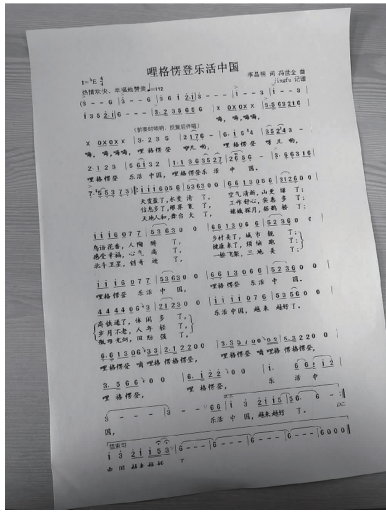
处理得到的最优分割,与图 6(b) 相比较,并未出现过分割的情况,同时尽可能地保留了简谱图像的音符和歌词信息,从而为后续的处理奠定了基础.

2.2 基于 DNN 的 PCNN 最优分割迭代次数的确定

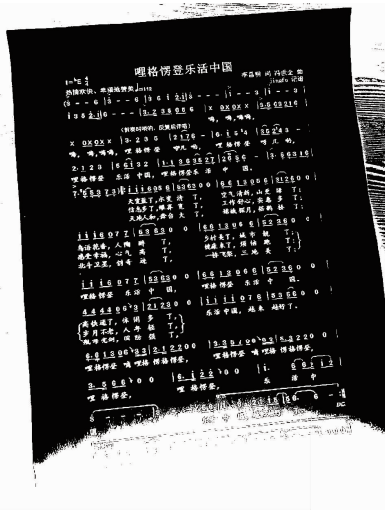
通过大量实验发现,PCNN 的迭代输出中哪一次为最优分割效果,与简谱图像本身的灰度统计分布有关,相同的灰度分布在 PCNN 迭代输出中具有相同的最优迭代次数. 但考虑到人工真实拍摄的简谱图,每一幅图像受光照影响都具有不同的灰度统计特性,无法用常规的方法对所有情况归纳出一般规律,所以对于这种复杂情况下的映射关系问题,利用 DNN 网络的学习能力是一种较好的解决办法. 为实现简谱图像灰度统计特性与 PCNN 最优分割迭代次数之间关系的映射,所构造的 DNN 网络结构如表 1 所示.

表 1 DNN 网络结构

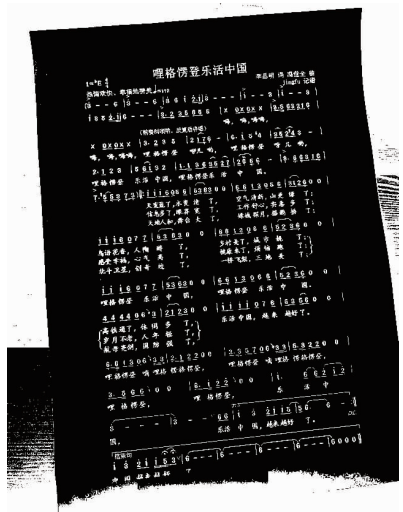
Layer type and network function	Parameters
Input	256 nodes
Hidden 1	17 nodes
Hidden 2	13 nodes
Output	1 nodes
Activation function of hidden	logsig
Activation function of output	logsig
Training function	trainlm
Training samples	540
Normalization of output	[0,255] to [0,1]



(a)人工拍摄图



(b)整幅图像的PCNN分割



(c)分块图像的PCNN分割

图6 分割结果比较

我们采用以下方案来训练 DNN 网络:(1)音乐网站下载标准简谱图像 20 张,每幅图像分成 9 个分块图像;(2)人工真实拍摄图像 50 张,每幅图像进行自适应分块处理. 共计得到 540 张的分块图像作为训练样本,然后通过人工方式对每一张图像的 PCNN 最优分割迭代次数进行手工标定,DNN 训练完成后,便可利用新输入分块图像的灰度统计特性进行最优分割图的预测.

3 基于双尺度下降法的简谱图像倾斜校正

对于通过上述方法得到的简谱最优分割图,在人工真实拍摄的条件不可避免存在图像倾斜的情况,这样便会对后续简谱信息的提取造成困难,因此需要进一步对简谱图像进行倾斜校正处理. 传统方法是用 Hough 变换来对简谱图像进行倾斜校正,但当分割图像中没有清晰连续的线条时,该方法就会存在一定的局限性. 对简谱图像观察发现可得,简谱图像中均存在许多的小节线,当图像本身处于水平位置时,这些小节线的水平投影值可以达到最小. 根据这一规律,本文提出一种基于小节线水平投影的双尺度下降法,从而实现简谱最优分割图的倾斜校正.

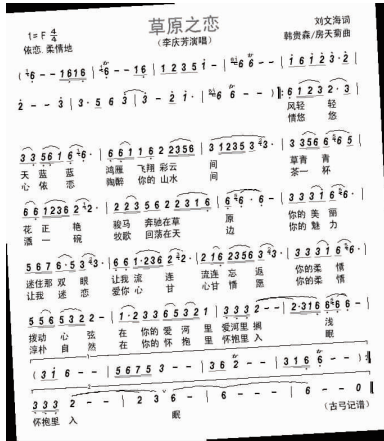
首先完成小节线的提取:根据简谱图像中小节线的长宽比(1/24),从最优分割图所有连通域中挑选出满足条件的连通域,再根据所得连通域长宽比的均值进一步滤除不满足该均值的连通域,从而得到仅保留有小节线(或部分)的分割图像,具体过程如图 7 所示.

图 7(a)为标准的简谱图像,经过 PCNN + DNN 处理后,在最优分割图的基础上进行连通域滤除,得到粗提取图 7(b)和精提取图 7(c). 可见图 7(c)在图 7(b)的基础上,清除了多余的干扰连通域,仅保留图像的小

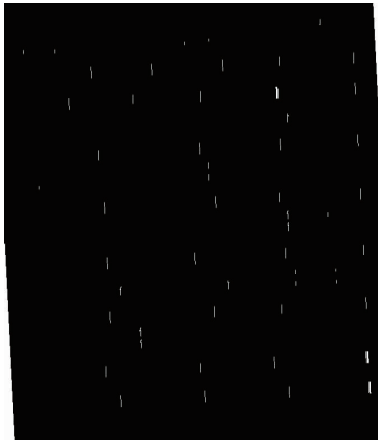
节线. 图 7(d)为人工拍摄的简谱图像,图 7(e)为粗提取,可见提取图中存在些许非小节线的多余连通域,但图 7(f)精提取图上滤除了这些多余的干扰.

第二步完成小节线精提取图的水平投影:经过大量的实验发现,当简谱图像水平时小节线的水平投影值(具体为水平投影二值化后再求和)最小. 当倾斜角度步长分别为 0.1° 和 1° 时,可以得到小节线水平投影值如图 8 所示.

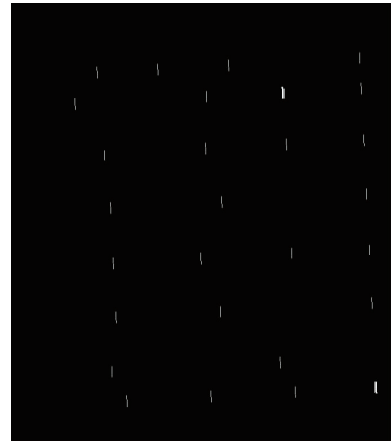
图 8(a)是标准简谱图像在不倾斜的情况下,图像以 1° 为单位,左转 15° 到右转 15° 时小节线的水平投影值. 可知,当简谱图像保持正中时,水平投影值最小,同时对称旋转角度之间的投影值大小基本相等,对于同一倾斜方向,简谱图像越趋于水平,其所对应的小节线的水平投影值就越小,即可以等效为一种梯度的下降,利用该特性便可实现简谱图像的倾斜校正,但明显存在校正精度不够的问题;图 8(b)是步长为 0.1° 时的小节线水平投影值,会发现在图像趋于水平的过程中,水平投影值并没有单调下降,在红色标记范围内清晰可见水平投影值出现跳变. 可见,仅仅通过减小步长的方式是无法判断倾斜方向的,进而也无法实现简谱图像的倾斜校正;图 8(c)是在区间 -1° 到 1° 范围,步长为 0.1° 时小节线的水平投影值,可见在同一倾斜方向又出现了单调下降的特性. 基于以上现象,为实现精细角度的倾斜校正,我们采用两种尺度的倾斜判断,先按 1° 为步长进行粗校正,当校正到 $\pm 1^\circ$ 时,再采用步长为 0.1° 的精校正. 由于在校正过程中采用了两种倾斜尺度,并且都具有梯度下降的特征,故我们称这种倾斜校正方法为双尺度下降法,利用该方法可以实现精度为 0.1° 的简谱图像自动倾斜校正. 具体过程如下.



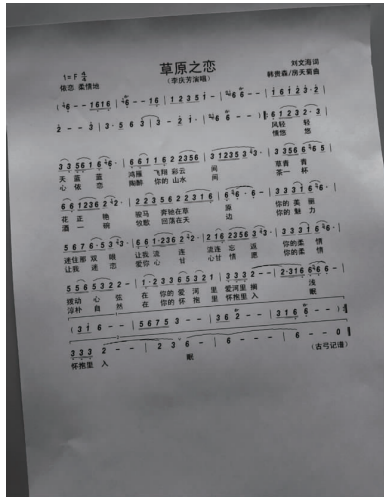
(a)标准简谱原图



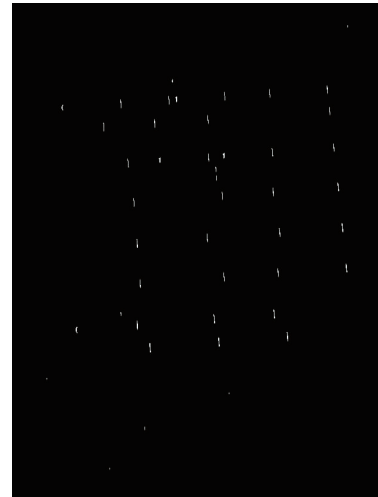
(b)小节线粗提取图



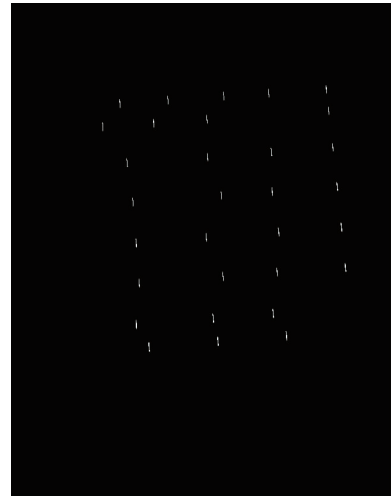
(c)小节线精提取图



(d)人工拍摄简谱图



(e)小节线粗提取图



(f)小节线精提取图

图7 简谱图像小节线提取图

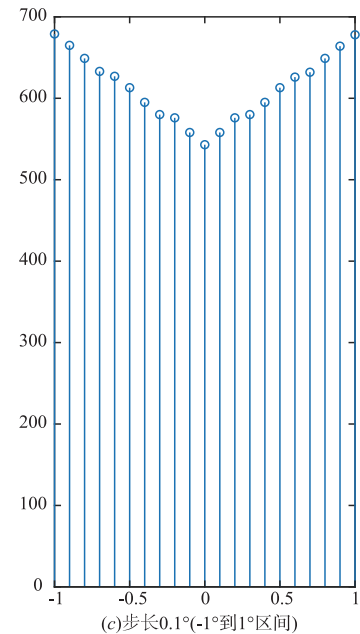
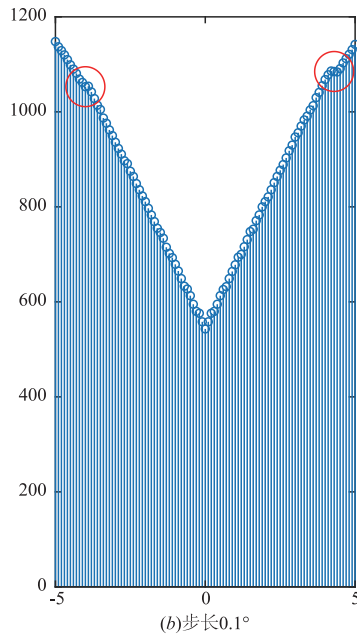
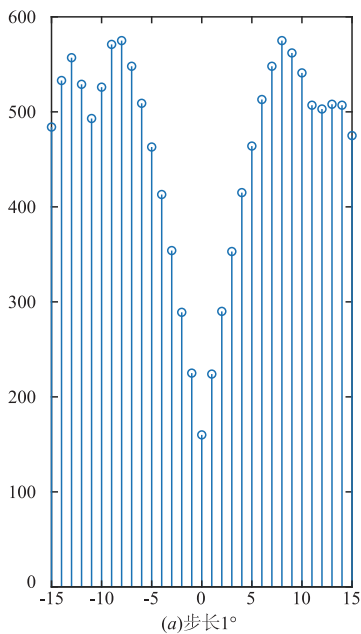
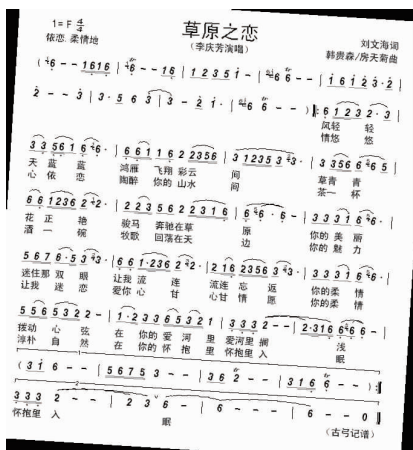


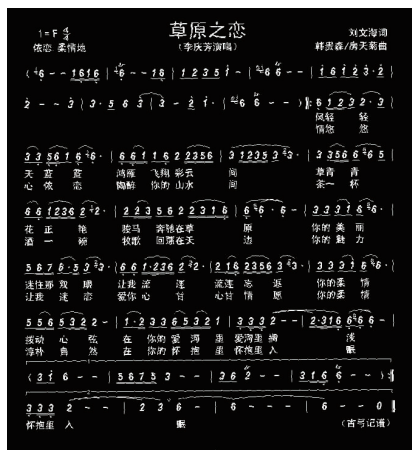
图8 不同步长旋转角度时小节线的水平投影值

第一步用 1° 的尺度来对图像进行粗校正. 首先, 不管图像往哪个方向倾斜, 第一次尺度调整均向右旋转图像 1° . 若调整后的水平投影值比调整前的值小, 说明原图像向右倾斜即调整方向正确, 继续以 1° 为步长向右旋转图像; 若调整后的值比调整前的值大, 说明原图像本身向右倾斜即调整方向错误, 此时则反方向以 1° 为步长向左调整; 当出现调整值比前一次调整值大时, 说明校正角度已经过了水平投影粗调整极小值对应的角度, 再依反方向旋转图像 1° , 便实现了简谱图像的粗校正, 此时简谱图像的倾斜角度已经落在 -1° 到 1° 区间. 第二步更改旋转角度为 0.1° 进行倾斜的精校正, 其处理步骤与粗校正相同. 具体校正过程如图 9 所示.

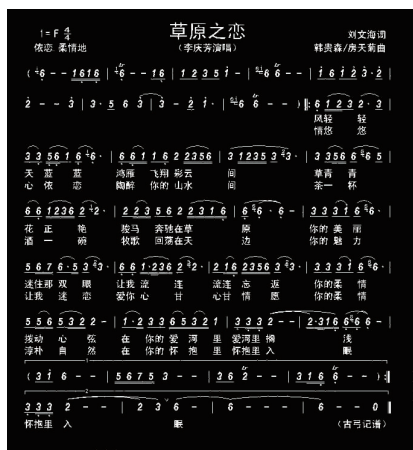
图 9(a) 是人为向右旋转 3.6° 的标准简谱图像, 图 9(b) 则是经过粗校正的图像, 图像校正为 -4° , 由于转动角度已知, 可见经过粗调整无法得到完全水平的简谱图像, 图 9(c) 则是最终精校正的结果, 可见与人为转动角度的大小相同方向相反, 实现的图像的倾斜校正. 图 9(d) 是一张人为拍摄的倾斜简谱图像, 人为测量倾斜角度大约为左转 4.5° , 图 9(e) 是图像的粗校正, 可见并未达到完全校正, 图 9(f) 经过精校正后旋转了 4.4° , 与人为测量的旋转角度基本接近, 可见本文提出的双尺度下降法不仅满足标准的简谱图像, 对于人工拍摄的简谱图像也能做到较为精确的倾斜校正.



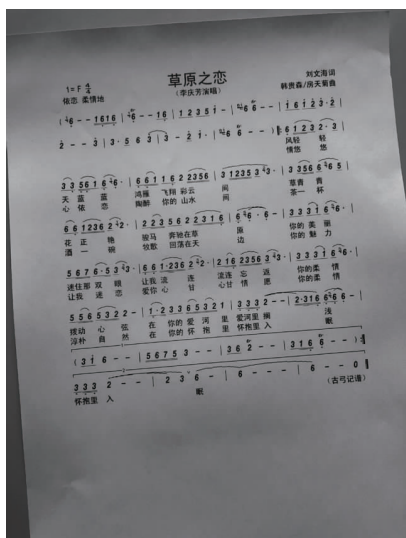
(a) 标准简谱未校正图
(已知旋转 3.6°)



(b) 标准简谱粗校正
(粗校正 -4°)



(c) 标准简谱精校正
(精校正 -3.6°)



(d) 人工拍摄未校正图
(人工测量倾斜角度 -4.5° 附近)



(e) 拍摄简谱粗校正
(粗校正 4°)



(f) 拍摄简谱精校正
(精校正 4.4°)

图9 标准简谱图像与人工拍摄简谱图像倾斜校正

4 简谱分割图中乐谱信息的提取

一张完整的简谱图像中包括: 音符, 歌词及标题.

通过对简谱图像排版结构的分析发现, 简谱小节线水平位置之间的部分为歌词部分, 同时第一个小节线之前的部分则为标题部分. 因此, 只要先根据前面校正图

像得到小节线的水平投影,其次再对整幅图像进行水平投影,便可清晰的区分音符部分和歌词部分.对倾斜校正过的图 9(c) 进行水平投影,其水平投影如图 10 所示.

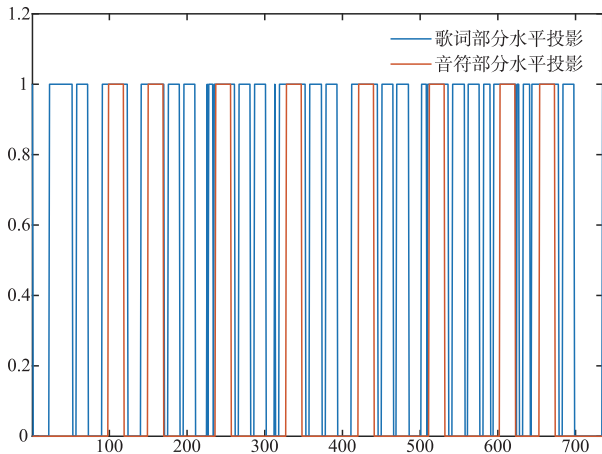
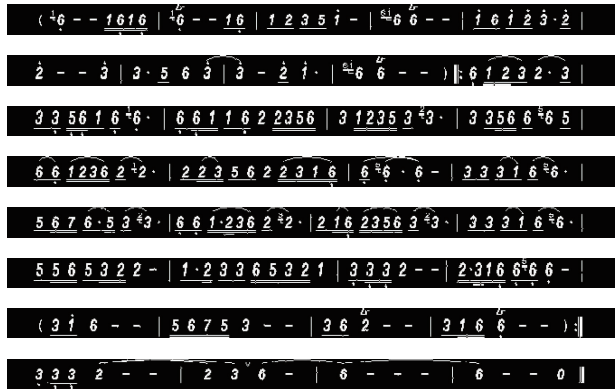


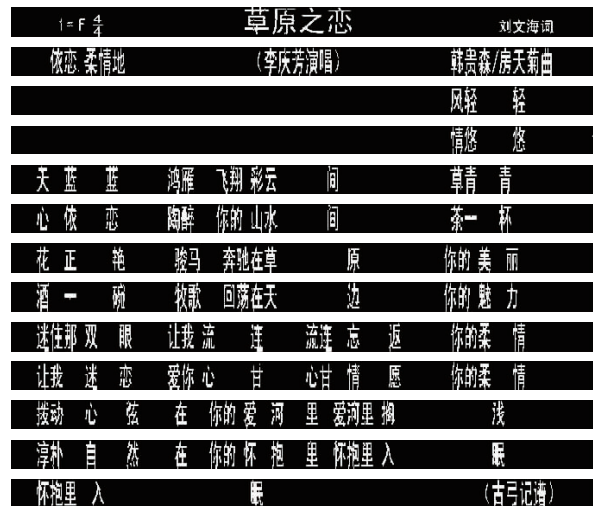
图10 标准简谱分割图水平投影



(a) 简谱提取及规则化

如图 10 所示,可清晰地看出图像的水平投影值被区分为红色的音符部分和蓝色的歌词部分,因此可直接根据这两个部分,对整幅简谱图像中的音符和歌词分别进行提取.为满足后续识别的需要,还需进一步对提取出的内容进行长宽比规则化处理.

图 11(a) 是简谱图像中简谱部分的提取,且对提取的每一行简谱都进行了规则化处理,方便后续每一个音符的识别;图 11(b) 则是歌词部分的提取,同样可见每一行歌词都完整提取且进行了规则化处理.但是提取出的每行简谱中存在大量的减时线、连音线等,其会为简谱中音符的提取及识别带来不便,因此提取出的简谱行图像不能简单的直接做垂直投影来提取每个音符.如果直接做垂直投影处理,会使多余的连通域对音符的垂直位置造成干扰,从而造成垂直定位过宽或者相邻音符被连通的错误.因此,本文先对行简谱图做上下去边操作,然后对去边的行提取图做垂直投影,其次再依据去边图的垂直投影对行提取图中的每个音符进行提取.



(b) 歌词提取及规则化

图11 简谱图像乐谱信息的提取

图 12(a) 是图 11(a) 中第一行上下去边之后的结果,可见,去除了连音符和符尾等的影响,将此图像进行垂直投影便可排除多余连通域的干扰.其中绿色线代表连通域的起点,紫色线代表连通域的终点,每个起终点的垂直区域代表一个音符.图 12(b) 则是以图 12(a) 为基准,在已经标明音符起终点的情况下对整行图进行的垂直投影,可见,每一个区域中仅包括音符及音节.图 12(c) 是在图 12(b) 的基础上对每一个标明的连通域进行提取,即提取分割出来的每一个音符,可见所有音符均提取完整且正确.

歌词部分的提取与音符提取方法不同,由于汉字中有偏旁的存在,若对其直接做垂直投影处理,会将偏

旁单独提取出来为后续带来不便,因此提取出的歌词行图像不能简单的直接做垂直投影来提取每个汉字.本文先对歌词第一行图像中的连通域从左到右标记个数,歌词提取算法可表示为式(10)~(14)所示:

$$r_i^{\min-l} = a_i \tag{10}$$

$$l_i = (r_{i+1}^{\min-l} - r_i^{\min-l}) - (r_{i+2}^{\min-l} - r_{i+1}^{\min-l}) \tag{11}$$

$$r_i^l = r_i^{\max-l} - r_i^{\min-l} \tag{12}$$

$$s_i = a_{i+1} - a_i \tag{13}$$

$$r_{\max}^l = \max(r_i^l) \tag{14}$$

其中, i 表示歌词行提取图中的第 i 个连通域, $r_i^{\min-l}$ 表示第 i 个连通域的最小列值为 a_i ,为了考虑括号对汉字的影响,计算任何一个连通域 i 的宽即 r_i^l ,由于简谱图

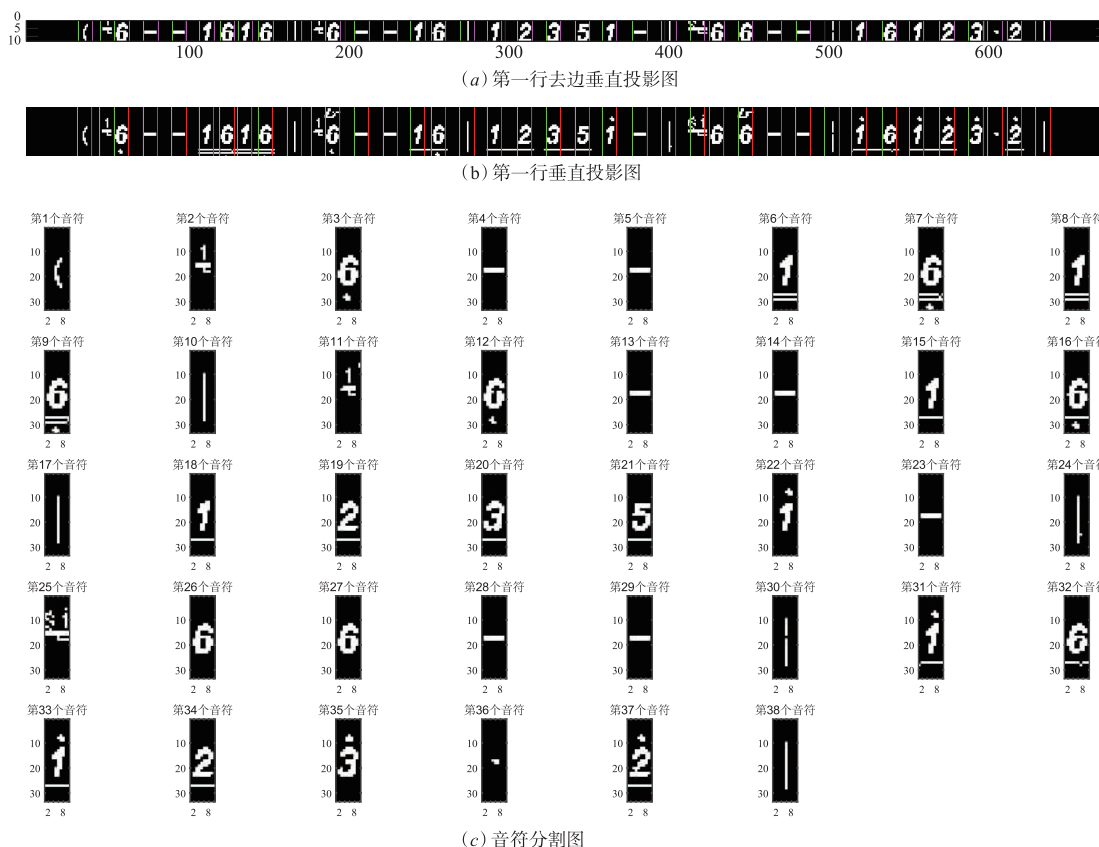


图12 简谱音符分割图

像仅在歌词最后存在括号,因此 r_{max}^l 表示之前确定的汉字连通域的最大宽, s_i 是相邻两个连通域最小行的差值, l_i 则是第 i 个连通域与第 $i+1$ 连通域之间的距离和第 $i+1$ 个连通域与第 $i+2$ 连通域之间的距离差值,即第 $i+1$ 连通域到前后两个连通域距离的差值. 若 $(r_{i+1}^{min-l} - r_i^{min-l}) < (r_{i+2}^{min-l} - r_{i+1}^{min-l})$, $s_i < \max(r_{i+1}^l, r_{i+2}^l)$ 且连通域均小于 r_{max}^l 时,则第 i 连通域与第 $i+1$ 连通域共为一个汉字,若不满足 $s_i < \max(r_{i+1}^l, r_{i+2}^l)$ 条件或任意一

个连通域相似于 r_{max}^l ,则三个连通域均为独立的汉字. 反之亦然. 若 $(r_{i+1}^{min-l} - r_i^{min-l}) \approx (r_{i+2}^{min-l} - r_{i+1}^{min-l})$,即三个连通域之间的距离近似相等,则三个连通域均为独立的汉字个体. 简谱图像歌词部分的提取如图 13 所示.

图 13 则是以图 11(b) 中的第一行为例,以上述方法得到的歌词分割图,可见每一个字符分割正确且完整,没有出现偏旁与汉字分开的情况,从而为后续的认识打下基础.

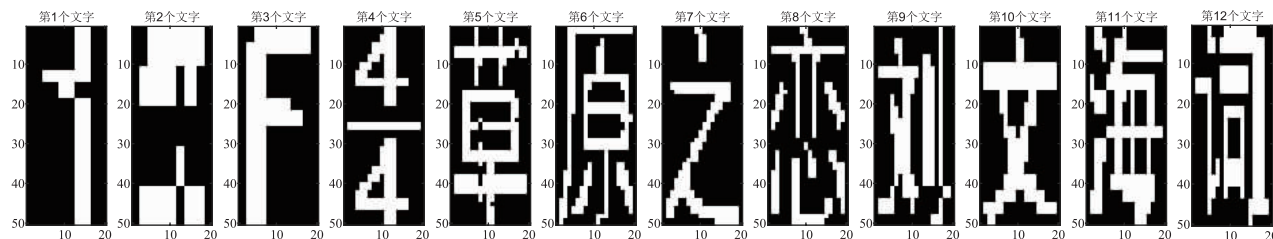


图13 简谱汉字分割图

5 算法流程及实验仿真分析

针对于纸质简谱图像的处理,本文主要完成了分割、倾斜校正、简谱信息提取三方面的工作,具体算法流程如图 14 所示.

在整个简谱图像的数字化的过程中,分割质量的好坏直接影响了后续环节的处理效果,为表明本文提出的分块 PCNN + DNN 图像分割算法的性能,我们将本文算法与其他几种经典的图像分割算法进行了比较,分割效果如图 15 所示.

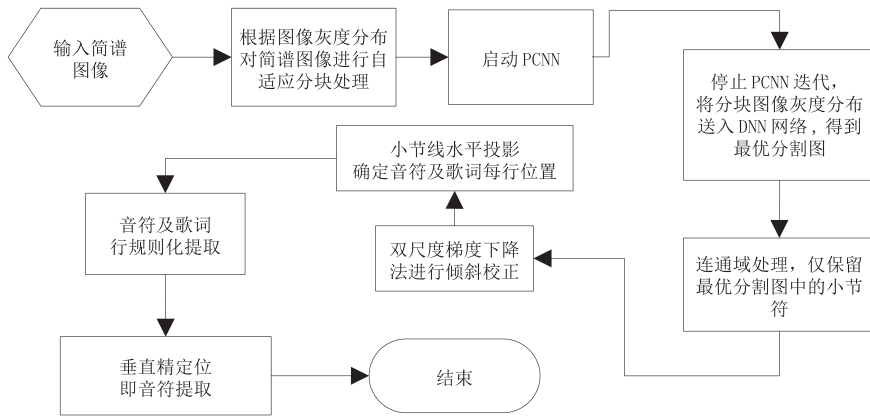


图14 简谱图像信息提取算法流程

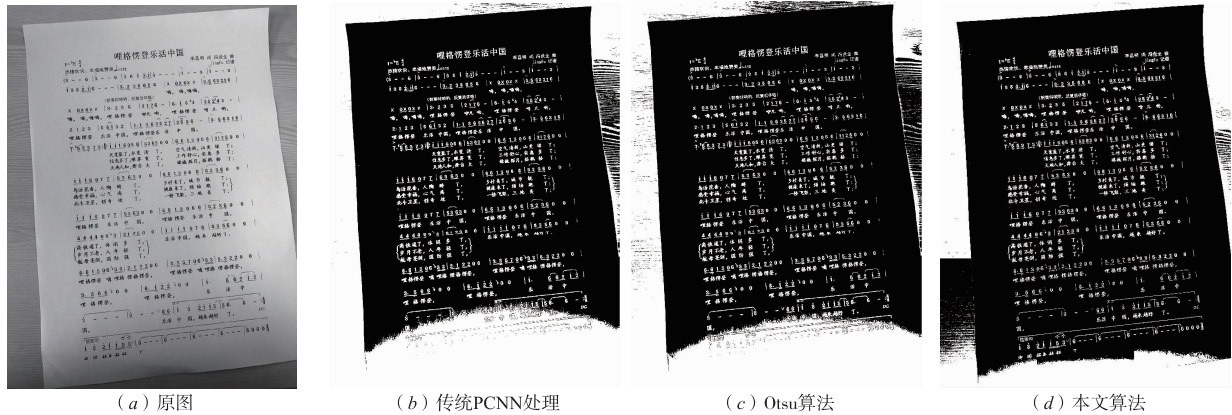


图15 与其他算法的比较

图 15(b) 和(c) 为两种经典图像分割算法的效果, 简谱图的下方均出现过分割情况. 图 15(d) 为本文算法的分割效果, 可见本文算法在处理人工拍摄简谱图时全局分割效果更好, 同时也保证了更好的简谱信息细节, 因此本文算法更适用于人工拍摄的简谱图像分割处理.

为验证本文所提出算法对简谱图像倾斜校正的鲁棒性和有效性, 实验用简谱图采用 50 张标准简谱图像和 50 张人工拍摄简谱图像. 在主频 3.6GHz、内存 8G 的计算机上使用 MATLAB R2018b 平台对算法进行仿真, 设计了不同分辨率以及不同倾斜角度情况下的实验. 表 2 为简谱图像分辨率为 800×600 时, 本文算法与传统的简谱倾斜校正算法的性能比较, 其中正确率是指当图像精校正之后与原始图像倾斜误差在 $\pm 0.1^\circ$ 时, 视为校正正确. 表 3 为本文算法在不同分辨率下, 不同倾斜角度简谱图像的分割及简谱信息提取效果.

表 2 简谱倾斜校正常见算法的性能比较

方法	图片数量/张	平均消耗时间/s	正确率/%
Radon 变换	100	4.47	72
Hough 变换	100	1.06	86
本文算法	100	0.39	98

表 3 不同分辨率且不同倾斜角度下的算法性能

分辨率	简谱倾斜角度	简谱倾斜校正平均消耗时间/s	单字符平均消耗时间/s
800×600	-6.4	0.37	0.143
	-4.3	0.32	
	0.0	0.11	
	5.2	0.35	
640×480	7.0	0.54	0.139
	-8.0	0.43	
	-3.6	0.27	
	0.0	0.10	
	3.0	0.24	
	5.7	0.31	

由表 2 可见, 传统 Hough 变换算法的运算速度虽然较快, 但仅当简谱中存在乐谱“长连线”时才可使用, 对于乐谱中没有“长连线”的简谱图像具有局限性. Radon 变换是将图像矩阵在某一指定角度射线方向上做投影变换, 因此对于标准下载图倾斜校正效果较好, 但对于真实拍摄的简谱图无法达到较好的效果. 是因为人为

拍摄过程难免出现由于拍摄角度不水平引起的简谱图像边缘形变的问题,进而导致简谱图像虽然整体调正了,但里面的简谱信息仍然存在一定程度的倾斜.而本文算法较其他两种算法更具鲁棒性和时效性,不仅仅对标准简谱图像能进行倾斜校正,对于人工拍摄的简谱图像也能做到准确的倾斜校正.

由表 3 可见,本文倾斜校正算法不受分辨率的影响,在不同分辨率条件下,均能保持较高的正确率和较快的处理速度.当旋转角度超过 $\pm 8^\circ$ 时,本文算法正确率明显降低,这是由小节线水平投影值单调下降特性的消失所引起的,但本文算法对 $-8^\circ \sim +8^\circ$ 范围内倾斜的简谱图像均有较好的校正效果,而且音符分割时间与倾斜角度基本无关,消耗时间很小.

6 结论

传统 PCNN 在用于全局灰度分布不均匀的图像处理时,往往会出现分割不均匀的情况.本文利用对全局图像分块处理的方式,分析分块模式及数量与最优分割之间的关系,从而提出了一种基于 PCNN 的简谱分块图像最优分割算法.在用于简谱图像分割时,该模型仅需利用 DNN 网络预测便可实现最优分割图的自适应选择.在得到简谱分割图之后,依据简谱图像小节线特点,发现简谱图像中倾斜角度与小节线的水平投影值有关,在倾斜角度大尺度步长变化时其值具有对称单调性,而在精细步长变化时其值仅在简谱图像的水平位置附近具有单调性.为此,我们提出了一种双尺度下降法来实现倾斜简谱的水平校正.该方法首先通过大尺度梯度下降法实现了简谱的粗校正,然后再利用小尺度梯度下降法进行精细调整,从而克服了单一尺度时要么校正精度不够要么无法校正的问题.在此基础上,根据音符区域和文字区域分别与小节线水平投影值之间的关系,分析了简谱图像的排版结构和不同区域的水平投影特征信息,运用所提出的去边垂直投影法和连通域距离判断法对两种区域分别进行处理,进而实现两种区域中的每个音符和文字的准确提取.实验仿真表明,本文算法对于各种复杂光照条件下拍摄的简谱图像都具有较好的鲁棒性,但在倾斜校正上还无法完成简谱图像大角度倾斜时的校正,另外,本文对提取出来的简谱及歌词信息没有进行识别处理,这些问题都是下一步研究中需要继续完成的工作.

参考文献

- [1] 刘晓翔,张树生,王贺,等.计算机光学乐谱识别技术[J].计算机工程,2003,29(2):14-15,27.
- [2] Calvo-Zaragoza J, Gallego A J, Pertusa A, et al. Recognition of handwritten music symbols with convolutional neural codes[A]. International Conference on Document Analysis and Recognition [C]. NY, USA: ICDAR, 2017. 691-696.
- [3] Calvozaragoza J, Toselli A H, Vidal E, et al. Handwritten music recognition for mensural notation with convolutional recurrent neural networks[J]. Pattern Recognition Letters, 2019, 128: 115-121.
- [4] Calvozaragoza J, Toselli A H, Vidal E, et al. Hybrid hidden Markov models and artificial neural networks for handwritten music recognition in mensural notation[J]. Pattern Analysis and Applications, 2019, 22(4): 1573-1584.
- [5] 陈根方,张文俊.基于 OMR 技术的乐谱图像匹配算法[J].计算机工程,2007,33(07):169-171.
- [6] CHEN Gen-fang, ZHANG Wen-jun. Music image matching algorithm based on OMR technique[J]. Computer Engineering, 2007, 33(07): 169-171. (in Chinese)
- [7] Reed K T, Parker J R. Automatic computer recognition of printed music [A]. International Conference on Pattern Recognition [C]. NY, USA: IEEE, 1996. 803-807.
- [8] Baro A, Riba P, Calvozaragoza J, et al. From optical music recognition to handwritten music recognition: a baseline [J]. Pattern Recognition Letters, 2019, 123: 1-8.
- [9] 程立,姚为,李波.基于投影轮廓的文本图像倾斜检测[J].中国图象图形学报,2015,20(01):29-38.
- [10] CHENG Li, YAO Wei, LI Bo. Improved projection profile based algorithm for skew detecting in document images [J]. Journal of Image and Graphics, 2015, 20(01): 29-38. (in Chinese)
- [11] Tigang Jiang, Fu qi, Yuming Mao. Numbered musical notation image recognition algorithm and realization [A]. The 9th Biennial Conference of the Australian Pattern Recognition Society [C]. NY, USA: DICTA, 2007. 274-277.
- [12] 高强.音乐教学中谱音识别的研究与实现[D].湖北武汉:武汉工程大学,2015.
- [13] GAO Qiang. The Research and Implementation of Staff and Music Identification in Music Teaching [D]. Wuhan, Hubei: Wuhan Institute of Technology, 2015. (in Chinese)
- [14] Rebelo A, Capela G, Cardoso J S. Optical recognition of music symbols[J]. International Journal on Document Analysis & Recognition, 2010, 13(1): 19-31.
- [15] Luth N. Automatic identification of music notations [A]. Proceedings of the Second International Conference on Web Delivering of Music (WEDELMUSIC 2002) [C]. NY, USA: IEEE, 2002. 203-210

- [13] 刘晓翔. 乐谱图像乐符分割技术[J]. 计算机应用研究, 2010,27(02):784-787,790.
LIU Xiao-xiang. Music symbols segmentation in music score images [J]. Application Research of Computers, 2010,27(02):784-787,790. (in Chinese)
- [14] Deng Xiangyu, Ma Yide. PCNN model analysis and its automatic parameters determination in image segmentation and edge detection[J]. Chinese Journal of Electronics, 2014,23(1):97-103.
- [15] Xiangyu Deng, Wenjuan Qin, Ran Zhang, et al. Automatic segmentation algorithm of license plate image based on PCNN and DNN[A]. Proceedings of International Conference on Image and Video Processing, and Artificial Intelligence[C]. NY, USA: SPIE, 2019. 1132102(1-9).
- [16] 邓翔宇, 马义德. PCNN 参数自适应设定及其模型的改进[J]. 电子学报, 2012,40(5):955-964.
DENG Xiang-yu, MA Yi-de. PCNN model automatic parameters determination and its modified model[J]. Acta Electronica Sinica, 2012,40(5):955-964. (in Chinese)
- [17] Eckhorn R, Reitboeck H J, Arndt M, et al. Feature linking via synchronization among distributed assemblies; Simulation of results from cat visual cortex[J]. Neural Computation, 1990,2(3):293-307.
- [18] Johnson J L, Padgett M L. PCNN models and applications [J]. IEEE Transactions on Neural Networks, 1999,10(3):480-498.
- [19] Xiangyu Deng, Yide Ma, Ming Dong. A new adaptive filtering method for removing salt and pepper noise based on multilayered PCNN [J]. Pattern Recognition Letters, 2016,79:8-17.
- [20] Dongguo Zhou, Yanhua Shao. Region growing for image segmentation using an extended PCNN model [J]. IET Image Processing, 2018,12(5):729-737.
- [21] Wei S, Hong Q, Hou M. Automatic image segmentation based on PCNN with adaptive threshold time constant [J]. Neurocomputing, 2011,74(9):1485-1491.
- [22] 毕英伟, 邱天爽. 一种基于简化 PCNN 的自适应图像分割方法[J]. 电子学报, 2005,33(4):647-650.
BI Ying-wei, QIU Tian-shuang. An adaptive image segmentation method based on a simplified PCNN [J]. Acta Electronica Sinica, 2005,33(4):647-650. (in Chinese)
- [23] Yang Z, Dong M, Guo Y, et al. A new method of microcalcifications detection in digitized mammograms based on improved Simplified PCNN [J]. Neurocomputing, 2016,218:79-90.
- [24] Xiangyu Deng, Chunman Yan, Yide Ma. PCNN Mechanism and Its Parameter Settings [J]. IEEE Transactions on Neural Networks and Learning Systems, 2020,31(2):488-501.

作者简介



邓翔宇(通讯作者) 男,1974年10月出生,甘肃康县人. 博士、教授、硕士生导师. 主要研究方向为人工智能、数字图像处理等.
E-mail:dengxy000@126.com



杨雅涵 女,1997年3月出生,甘肃兰州人. 现为西北师范大学电子信息专业硕士研究生. 主要研究领域为神经网络及其在图像处理中的应用.
E-mail:yangyh111@126.com