

基于视觉的错序在架图书检测系统

康洪雷¹, 牛连强¹, 冯庸², 秦月¹, 鄂雪娇¹, 刘宇航¹

(1.沈阳工业大学软件学院, 辽宁 沈阳, 110023;

2.沈阳工业大学理学院, 辽宁 沈阳, 110023)

摘要:在架图书整理是图书馆的一项常见工作,迫切需要提高自动化水平。本文以尽可能不增加额外投资和设施改造为前提,提出了依靠移动设备获取图像,通过视觉技术构建软件系统以实现对错序图书的自动识别的方法。该方法由移动设备采集在架图书图像,利用LSD算法检测书脊边缘并分割图书图像,进而得到索书号区域,并通过投影分离出索书号区域中的字符。采用前馈连续双卷积CNN网络结构模型并综合使用了Dropout、ReLU激活函数和Softmax回归等技术完成对字符的识别。最后,将组合后的索书号与数据库预存信息进行比较,得到图书的正确排列信息。对系统的实际测试和分析表明,本文的方法可以针对一般索书号在架图书进行较好的错序检测,能为图书的自动化整理提供一种可行的方案。

关键词:错序图书;边缘检测;直线检测;卷积神经网络;索书号

中图分类号: TP301 **文献标识码:** A

A Vision-Based System to Detect Books with Incorrect Sequence on Shelf

KANG Honglei¹, NIU Lianqiang¹, FENG Yong², QIN Yue¹, E Xuejiao¹, LIU Yuhang¹

(1. Software School, Shenyang University of Technology School, Shenyang 110023, China;

2. Science School, Shenyang University of Technology, Shenyang 110023, China)

Abstract: Sorting books is a regular task of libraries, and it's urgent to improve the automation level. Without any additional investment and infrastructure improvement, an approach to detect books with incorrect sequence on shelf by means of a software system based on vision inspection technique is proposed. Firstly, mobile devices are employed to capture images of books on shelf. Secondly, LSD algorithm is used to find all edge lines of book spines, and characters of call number are able to be gained by means of projection from ROI separated from a book image. Thirdly, a feed forward continuous double convolution network model is built to perform the task of character recognition, where other techniques such as Dropout, ReLU activation function and Softmax regression are also included. Finally, the reconstituted call numbers are compared with the recorded information in a database to obtain correct positions. Experiments and analysis by running the system show that the approach is able to ensure good performance on finding those books with incorrect sequence on the shelf.

Keywords: books with incorrect sequence; edge detection; line detection; CNN (Convolutional Neural Network); call number

1 引言(Introduction)

实现对图书的自动识别、检索、整理有着广泛的应用前景,典型地,可有效地解决一种常见的应用——乱架图书的整理工作。乱架图书整理是图书馆日常维护工作的一个重要组成部分,是一项循环、耗时、费力且易于出错的任务,目前主要依赖人工完成^[1]。为了实现对图书馆书籍的自动查架及整理,一些技术如条形码、二维码^[2,3]、无线射频(RFID)^[4,5]、机器人(机械手臂)被应用到在架图书盘存过程中。对该问题的研究涉及到对图书本身(包括书脊、书名和索书号等)的识别和系统构建两方面的问题,问题的解决可以使工作效率和经济

效益得到显著提高,因而受到广泛关注。

2009年, Lee D等^[6]通过CMOS相机采集书脊图像,利用直方图匹配方法确定图书是否摆放正确或丢失。王圆妹^[7]参考车牌识别方法,将CCD摄像头获取的书脊上索书号图像经一系列图像变换获取索书号字符图像,再通过模版匹配方法识别图书索书号。Chen D M等^[8]于2010年提出了一种通过智能手机拍摄图像,经Hough变换查找书脊位置,再提取segSURF特征进而识别书脊的方法,可以达到74.5%的准确率。Wing W Y NG等^[9]于2011年通过径向基函数网络粗定位包含目标RFID标签的图书,利用书脊图像的SIFT特征与数

数据库预存值进行比对, 以实现图书的智能定位。Fowers S等^[10]提出了基于彩色图像特征CDSIFT对图书进行识别的方法, 随后, Duan X等^[11]则提出了利用图书索书号来识别图书的思路及通过RGB空间定位索书号标签的方法。方建军、Nevetha M P利用索书号标签在HSV空间中的分布特征, 结合霍夫变换检测书脊边缘等方法实现了索书号提取^[12,13]。

总体上, 采用附加设备或技术涉及到对现有书库及存储系统的大规模改造, 需要大量人力和财力投入, 而利用移动设备和视觉识别进行图书识别可以以更小的代价实现对现有在架图书的维护。本文在前期工作^[14]的基础上, 以移动设备获取图像为出发点, 结合图像处理、直线检测和深度学习技术获取索书号, 再利用后台数据存储为依据进行错序图书判别, 从而构建起一种有效的在架图书错序检测系统。

2 系统构成与工作流程(Composition and process of system)

对于采用一般设备获取的在架图书的图像, 可以直接依赖视觉识别技术检测出书脊图像和索书号区域, 分割出索书号字符, 进而识别出这些字符。重新组合为索书号序列后, 通过网络与数据库交互, 进而判别其是否错序并给出反馈, 基本流程如图1所示。

从降低造价和易于实现两方面考虑, 采用移动设备(如手机)采集图像是一种可取的选择, 下述流程进一步说明了各节点的工作和处理技术。

(1)利用移动设备手工采集在架图书图像, 传至应用系统(服务器)。

(2)采用LSD直线提取算法在预处理后的图像中检测书脊直线, 对书脊定位。主要步骤包括:

Step1: 对采集图像进行图像预处理, 包括颜色变换、去噪、二值化等;

Step2: 结合图书书脊边缘特征, 利用LSD直线检测算法检测书脊边缘线;

Step3: 按一定规则对书脊边缘线进行归并、取舍, 得到真实的书脊线;

Step4: 依据书脊线确定的书脊, 以及索书号区域的特征, 分割出索书号区域;

Step5: 对索书号区域进行倾斜校正, 分割出索书号字符。

(3)采用CNN卷积神经网络实现对索书号字符的识别。

(4)将字符组合成索书号。

(5)通过检索图书索书号与数据库预存信息生成图书摆放正确与否的检测结果。系统将所有馆藏图书的索书号信息、摆放的书架号、行号、列号、状态等信息预先保存在数据库, 将实际识别出的图书索书号与数据库进行对比, 生成检索报告。

本文分别采用OpenCV 3.3、Caffe框架支持图像处理和文字识别, 并利用VS2013C++实现完整系统。

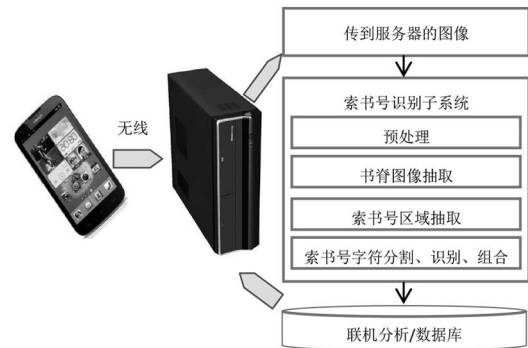


图1系统结构

Fig.1 System organization

3 图书索书号识别(Recognition of call number)

3.1 书脊检测

在图书馆中, 索书号是一种普遍采用的在架图书管理手段, 是查找、定位和排列书籍的基础。索书号通常位于书籍的下部, 一般需要先先将一整幅图书图像分割成单个图书图像, 而分割的基础是书脊边缘几乎是直线。

现有方法一般采用Canny算子获取图像边缘, 再利用Hough变换提取边缘图像中的直线。本文首先对图像做灰度转换:

```
void cvtColor(InputArray src, OutputArray dst, int code, int dstCn=0)
```

除了灰度外, 此方法也支持由RGB向HSV和HSI颜色空间转换。

利用createLineSegmentDetector()函数创建与初始化一个linesegmentdetector类对象:

```
P t r < L i n e S e g m e n t D e t e c t o r >
createLineSegmentDetector(int _refine = LSD_REFINE_STD, ...)
```

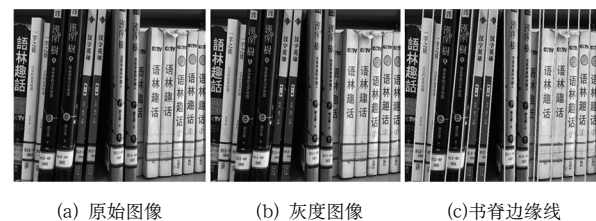
将参数_refine指定为LSD_REFINE_NONE, 其他参数采用缺省值, 调用此对象的detect()方法, 并将灰度图像作为方法的_image参数值:

```
void detect(InputArray _image, OutputArray _lines, ...)
```

利用参数_lines得到检测出的所有直线信息。

相较于Hough变换, LSD算法可以在线性时间内得出亚像素级精度的结果^[15], 且无需调节参数, 也不需要伪直线进行处理, 能够有效地避免阈值选择、书脊边缘线断裂和扭曲、书脊文本误检为直线等问题。

图2显示了一次书脊边缘线的检测结果。



(a) 原始图像

(b) 灰度图像

(c) 书脊边缘线

图2 书脊边缘线检测结果

Fig.2 Measuring result of edge of book spine

3.2 索书号区域提取

为了提高区域分割的准确性,首先依据以提取的书脊直线对每本图书图像进行倾斜校正,只需要将图像旋转一个角度,但可以有效避免现存方法中可能存在的后续区域提取不完整的问题。

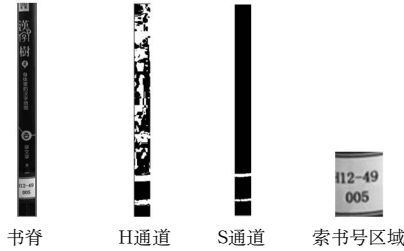


图3 索书号区域提取

Fig.3 Extracting region of call number

在分割单本图书图像后,索书号区域是下一步的感兴趣区域(ROI)。不过,索书号通常是图书馆自己设计的一种标签,外观、颜色、文字反向和粘贴位置等均不固定,只能依据具体情况设计针对性的提取策略。图2(a)显示了沈阳工业大学图书馆在架图书索书号的样例。通常,索书号标签有一个边款,索书号字符和背景颜色差别较,索书号宽高比有一定比例等特征,从而构成了分割该区域的基础。

考虑到颜色分量R、G、B间有较高相关性,本文采用主观彩色模型HSV颜色空间,以更好地反映人对颜色的感受。在HSV空间的三个参数中,H、S反应了色彩的本质特性,而V代表着亮度,故可以进使用H和S参数,以消除光线强度变化对书脊图像的影响。

为了确定索书号区域的范围,本文采用了文献[12]提出的对大量索书号区域的直方图统计方法。利用直方图确定区域对应的色调分量H和饱和度分量S的值域,确定H分量和S分量取值范围,进而得到索书号区域。不过,由于本文已先行将单本图书分开,不再需要后续的处理工作。图3显示了一次提取结果。

3.3 索书号字符分割

由于前文已经对每个单本图书图像进行了倾斜校正,故可以采用简单投影法进行索书号字符分割。在转换为灰度图像后,索书号字符通常为黑色或颜色较深,而背景接近白色,二者灰度差异较大,且索书号字符大小、排列较规则,且通常在两行以内。

于是,可以先确定三个阈值threshold、 Δt_1 和 Δt_2 ,并对图像进行水平投影以分割出字符所在行,方法是遍历灰度图像所有行,记录该行内灰度值小于threshold的点数 c 。若 $c \geq \Delta t_1$,则表示该行为字符所在行。在该上的两侧,会发生点数 c 由大到小的跳变,意味着文字行的上下两端边界。若连续两行的点数 c 之比满足 $c_{\min} / c_{\max} \leq \Delta t_2$,则认为属于字符行的边界。

对于不同的索书号,上述阈值可以通过预先采集多幅图像进行试验的方法来确定,也可以寻求更为智能的方法。本文针对全文采集的实际图书图像,经试验得出threshold为130, $\Delta t_1 = \Delta t_2 = 8\%$ 。

按类似的方法对文字行图像进行垂直投影即可得到最终

的分割字符。

图4给出了分割后的字符样例。

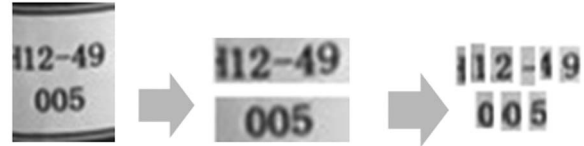


图4 字符分割

Fig.4 Character segmentation

3.4 基于CNN的索书号字符识别

自20世纪90年代以来,深度学习的模型结构逐渐已经占据模式识别领域的重要地位,尤其是卷积神经网络(CNN),在字符识别等领域获得了巨大成功。CNN是一种多层前馈深度学习模型,专门用以处理具有类似网格结构的数据。CNN以图像数据作为输入,以类别为输出,通过局部感知域提取图像特征,且通过权值共享在整个图像中提取相同特征,降低了计算开销。同时,利用下采样方法降低了特征图分辨率,保证了所提取的特征针对平移、缩放和形变的较好的不变性^[16-20]。

本文选用了一种前馈连续双卷积网络结构模型,它由一个输入层(索书号字符图像)、输出层(36类输出概率)、两个卷积层(5×5 大小卷积核)、两个全连接层组成,并引入Dropout集成方法提高网络泛化能力,采用ReLU激活函数缩短网络训练时间,其结构图如图5所示。

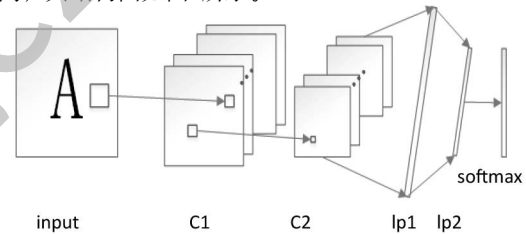


图5 前馈连续双卷积网络结构

Fig.5 A feedforward continuous double convolution network model

(1)输入层。卷积神经网络的输入层直接接受二维视觉模式,本文的网络输入为自然场景中采集的图书索书号字符图像,受网络规模、拍摄设备、字符分割等限制,需要将标签图像进行归一化处理,将分割后的单个字符统一缩放至 28×28 像素大小。为保持连通性与拓扑结构,采用了归一化方法。此外,为了提高识别准确率,对归一化后的图像数据集实施了减去均值图像的白化操作。

(2)卷积层。卷积层对应着特征的提取,每个神经元从上一层局部感受域得到突触输入,从而提取局部特征,而具体提取的特征体现为局部感受域的连接权重上,相对于一般的前馈型网络结构,这种连接方式可以极大地缩小网络参数规模。

对于一个输入的二维字符图像 I 和卷积核 $W_{m \times n}$,经对图像卷积后得到图像特征 S 为:

$$S(i, j) = I \otimes W_{m \times n} = \sum_{p=1}^m \sum_{q=1}^n I(i+p, j+q) W(p, q)$$

网络利用多个卷积层提取到多个不同特征用于最终分类。

(3)Dropout、激活函数与Softmax回归。在系统实现时,本文充分考虑和采用了目前流行的一些技术以克服早期CNN网络中存在的问题。其中,以Dropout作为避免过拟合的措施,利用ReLU函数作为激活函数,以克服在反向传播过程中存在的“梯度消失”局限,提高网络的收敛速度。同时,作为一个多维分类问题,本文还采用了将训练集拆分为多个子集的Softmax回归方法,以促进网络能够提高得到最优解的可能性。

4 系统构建与实验(System construction and experiment)

4.1 系统构建

(1)Caffe框架。系统构建的主要问题是数据库和Caffe框架的引入,限于篇幅,仅扼要说明Caffe框架(<https://github.com/microsoft/caffe>),主要过程为:

①解压到文件夹“caffe-master”。

②处理“Windows”文件夹中的VS2013的解决方案,更名CommonSettings.props.example为CommonSettings.props,并设置文件中的CpuOnlyBuild选项为true。选项UseCuDNN可依据采用CPU/GPU训练来调整。

③修改“Windows”文件夹下的Caffe.sln文件内的Debug为Release,调整工程属性页的“C/C++->常规->将警告视为错误调整”为否。

点击编译测试解决方案里的caffe工程,因为引用了libcaffe工程,所以两个工程都会编译。

④编译所有项目,创建VS2013工程如BookDetector,并将Caffe的C++接口文件“caffe\caffe-master\examples\cpp_classification”添加到工程中。

⑤引入第三方库。即以动态链接库方式引入nuget下载的第三方库。

⑥添加Caffe的libcaffe静态库。

(2)错序检测。由于一些特殊原因如图书过薄、遮挡或模糊等原因会导致部分索书号字符缺失,本文采用正则表达式对数据进行模糊查询处理,以确定当前书的位置。随后,检查是否满足条件“前索书号 $\leq$ 当前索书号 $\leq$ 后索书号”,若是则表明图书摆放位置正确。在程序实现时,还需要考虑编码使用的字符和索书号残缺带来的影响。

4.2 实验与结果分析

(1)数据集。本文输入数据均通过从沈阳工业大学图书馆采集在架图书图像,进而分割出索书号单个字符和符号的方式获得。符号分为36类,包括数字0—9、大写英文字母(不含W和Y),以及“.”和“-”。每类符号图像采集100个构成样本集,样例如图6所示。



图6 字符集部分字符实例

Fig.6 Character samples of data set

(2)训练与测试。以测试与训练集规模比例为1:3且无交叉方式训练网络,可以达到95.12%的测试集识别率。

分析测试集中识别错误的图片,增加部分残缺字符、笔画偏粗的字符“1”等样本,并在训练时采用Caffe的LMDB图像格式。当迭代至3700次时,测试集的识别率达到99.36%。

(3)运行。实际采集在架图书图像500幅,均含有1—3个错序,共870个错序。经系统检测,正确发现错序数799个,占比91.8%。分析错误的原因,主要来自于较薄的图书,此时的索书号被翻折出书脊,无法获得完整的索书号,系统会产生明显的误判和漏判。对于此类问题,可以考虑通过多行索书号等方式进行改造,或者重新制定相应规则以使索书号可以更好地容错。

5 结论(Conclusion)

本文提出了一个结合图书馆现有设施和图书管理方式的自动化错序图书自动化检测方案,并综合运用了图像处理、快速直线检测、深度学习等技术给出了完整的实现方法。实际构建系统并针对图书馆的应用实验表明,该方法能够准确地检测出连续摆放的图书并提取出索书号区域,进而以较高的准确率识别出索书号,通过与数据库的预存信息比较发现摆放错位的图书,是提高图书整理工作自动化的有效方法。

参考文献(References)

- [1] 朱明,胥伟岚.2003—2012年国外图书馆管理系统研究述评[J].图书情报工作,2013,57(9):129—134.
- [2] 唐晓国.条形码技术及其在图书馆的应用[J].情报理论与实践,2000,23(5):383—385.
- [3] 李健,杨京峰,张成昱.条形码/二维码技术在图书馆个性化服务中的应用与实现[J].现代图书情报技术,2012(2):82—86.
- [4] 韩凤勇.无线射频技术在图书馆领域的应用[J].图书馆工作者研究,2010(10):37—39.
- [5] 安结.图书馆领域中淘汰条形码的新技术—RFID[J].图书馆学研究,2004(9):9—11.
- [6] Lee D,Chang Y,Archibald J,et al.Matching book-spine images for library shelf-reading process automation[C].IEEE International Conference on Automation Science and Engineering,2008:738—743.
- [7] 王圆妹.图书索书号自动识别技术的研究与实现[J].计算机与信息技术,2009(Z2):4—6.
- [8] Chen D M,Tsai S S,Girod B,et al.Building book inventories using smart phones[C].ACM Multimedia,2010.
- [9] Wing W Y NG,Qiao Y S,Lin L,et al.Intelligent book positioning for library using RFID and book spine matching[C].2011 International Conference on Machine Learning and Cybernetics,2011:465—470.
- [10] Fowers S,Lee D.An effective color addition to feature detection and description for book spine image matching[J].ISRN Machine Vision,2011,2012:1—15.
- [11] Duan X,Zhao Q,Anwar S.Identifying books in library using line segment detector and contour clustering[C].Computing

- and Convergence Technology (ICCCT),2012:998–1003.
- [12] 方建军,赵强强.图书馆在架图书的索书号图像提取与分割[J].北京联合大学学报,2015,29(1):87–92.
- [13] Nevetha M P,Baskar A.Automatic book spine extraction and recognition for library inventory management[C].Third International Symposium on Women in Computing and Informatics (WCI '15),2015:44–48.
- [14] 冯庸.基于视觉的图书馆在架错序图书自动识别技术研究[D].沈阳:沈阳工业大学,2017.
- [15] Von Gioi R G,Jakubowicz J,Morel J M,et al.Lsd:a line segment detector[EB/OL].<http://dx.doi.org/10.5201/ipol.2012.gjmr-lsd>,2012–03–24.
- [16] Jaderberg M,Simonyan K,Vedaldi A,et al.Reading text in the wild with convolutional neural networks[J].Int.J.of Computer Vision,2015:1–20.
- [17] Bengio Y.Learning deep architectures for AI[J].Foundations and Trends in Machine Learning,2009,2(1):1–127.
- [18] Kessy A,Lewin A,Strimmer K.Optimal Whitening and Decorrelation[EB/OL].<http://arxiv.org/abs/1512.00809v4>,2016–09–18.
- [19] Krizhevsky A,Sutskever L,Hinton G E.ImageNet classification with deep convolutional neural networks[J].Neural Information Processing Systems,2012,60(2):1097–1105.
- [20] Hinton G E,Srivastava N,Krizhevsky A,et al.Improving neural networks by preventing co-adaptation of feature detectors[EB/OL].<https://arxiv.org/abs/1207.0580>,2012–07–3.

作者简介:

康洪雷(1996–),男,本科生.研究领域:软件工程,信息处理.
牛连强(1965–),男,博士,教授.研究领域:CG&CAD,图像处理,信息可视化及仿真.
冯庸(1992–),男,硕士生.研究领域:图像处理.
秦月(1996–),女,本科生.研究领域:软件工程.
鄂雪娇(1997–),女,本科生.研究领域:软件工程.
刘宇航(1996–),男,本科生.研究领域:软件工程.

(上接第51页)

区域A,比较矩形区域A和矩形区域O的大小,如果矩形区域A大于矩形区域O设定范围,则通知屏幕交互模块显示黑屏蒙层页,并且语音和文字提醒观看者退回到适当的距离处才能消除黑屏。如果检测到矩形区域A已经在可接受的范围内,则通知屏幕交互模块取消黑屏蒙层页和语音提醒。如果没有检测到人脸,则生成的矩形区域A和全屏同样大。

2.4 计时模块

在开启语音控制时,将提示用户设置允许连续观看的时长,如果用户设置了无限制,则不开启计时模块,如果设置了某个时长,则开启计时模块开始计时,一旦计时到达设定时长时,则通知屏幕交互模块出现黑屏蒙层页和眼保健操音乐提醒休息眼睛。

3 结论(Conclusion)

本设计的有益效果:提供一种有效监视儿童屏幕距离、观看屏幕时间的方法,让儿童在使用智能终端时必须在预设的距离之外、有效的时间内才能流畅观看屏幕,一旦违反这些原则将出现音乐并黑屏提醒屏幕,只要修正,则可以继续返回观看。采用语音命令控制和蒙层屏幕显示相结合的方式,更加贴近儿童的经验,让儿童在观看节目的时候可以根据屏幕上显示的文本远距离地发出操作指令,不但指令清晰,而且可以训练儿童识数识字的能力。更进一步地,提供自动识别屏幕背景色

和寻找各按钮点击区域块的方法,将适用于各类智能视频终端,观看者只需要安装语音控制软件便可以对现有的视频终端进行语音控制,不需要各视频终端开发商进行二次开发。

参考文献(References)

- [1] 张杰,潘欣裕.一种噪声混叠下的阵列语音采集系统[J].单片机与嵌入式系统应用,2017(12):28–31.
- [2] 张建华.基于深度学习的语音识别应用研究[D].北京:北京邮电大学,2015.
- [3] 郭晓宇,平西建,周林.基于连通区域矩阵的文本图像识别方法[J].信息工程大学学报,2012(6):329–333.
- [4] 贺文伟,张玉玲.一种自适应的基于中心偏置特性的矩形搜索方法[J].计算机系统应用,2011(7):214–216.
- [5] 杨少波.基于阵列图像采集结构的液晶屏缺陷检测算法研究[D].成都:电子科技大学,2016.
- [6] 马宇尘.获取屏幕截图的方法及客户端[P].中国专利:CN105677142A,2016–06–15.
- [7] 雷琴琴.面向像素的可视化技术研究[D].北京:北京交通大学,2007.

作者简介:

李清霞(1970–),女,硕士,讲师.研究领域:数据挖掘,图象处理.