

上海交通大学硕士学位论文

用于读书机器人的版面分析和阅读顺序确定算法研究

学 校： 上海交通大学

学 院： 电子信息与电气工程学院

系 科： 自动化

专 业： 模式识别与智能系统

班 级： B0803293

学 号： 1080329110

姓 名： 潘昱存

导 师： 赵群飞（教授）

上海交通大学电子信息与电气工程学院

二〇一〇年十二月三十日

**A Dissertation Submitted to Shanghai Jiao Tong University for the
Degree of Master**

**Document Layout Analysis and Reading Order
Determination for a Reading Robot**

Author: Pan Yucun

Specialty: Institute of Image Processing and
Pattern Recognition

Advisor: Prof. Zhao Qunfei

School of Electronics and Electric Engineering

Shanghai Jiao Tong University

Shanghai, P.R.China

December 30, 2010

上海交通大学

学位论文原创性声明

本人郑重声明：所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不包含任何其他个人或集体已经发表或撰写过的作品成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。本人完全意识到本声明的法律结果由本人承担。

学位论文作者签名：

日期： 年 月 日

上海交通大学学位论文版权使用授权书

本学位论文作者完全了解学校有关保留、使用学位论文的规定，同意学校保留并向国家有关部门或机构送交论文的复印件和电子版，允许论文被查阅和借阅。本人授权上海交通大学可以将本学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存和汇编本学位论文。

保密☐，在____年解密后适用本授权书。

本学位论文属于

不保密☐。

（请在以上方框内打“√”）

学位论文作者签名：

指导教师签名：

日期：

年 月 日

日期：

年 月 日

用于读书机器人的版面分析和阅读顺序确定算法研究

摘 要

读书机器人 JoyTon, 利用光机电一体化系统, 将纸介质信息数字化, 然后通过文本语音转换, 将视觉信息转换为听觉信息, 为盲人和视力障碍者提供人性化的阅读服务。仅依靠光学字符识别技术, 读书机器人只能正确地阅读极少数的文档。究其原因, 其中很重要的一方面是, 文档图像通常会包括图像、表格、文字、线框等要素, 并以各种版面规则进行排列, 这些要素和规则, 是简单的光学字符识别所不能理解的。

版面分析和理解是对图像内的文本、图像、表格、线框等要素的特征和位置、逻辑关系进行自动的识别、分割和理解的过程, 本文着重对文档图像的版面分析和阅读顺序确定算法进行研究, 具体包括以下几方面:

(1) 文档图像预处理

由于读书机器人 JoyTon 通过控制数码相机对翻书机构中的纸质文档直接拍照, 获取图像的环境因素多变, 文档图像往往包含背景, 并且受到光照不均匀、对比度拉伸、物体阴影的影响, 为版面分析和阅读顺序确定增加了难度, 所以需要进行一系列的图像预处理, 从而提取单页文档图像, 提升图像质量, 进行几何扭曲校正, 并且完成图像二值化和缩

小。

（2）文档图像的版面分析

首先扫描预处理获得的二值缩小图像，抽取所有的连通域，然后根据连通域的尺寸特征，区分直线、图表和文字。对于直线，留作阅读顺序确定的信息。对于图表，根据它们的纵横投影，可以区分为图像、表格和带框文本；带框文本在去框后，作为单独图像进行分析，获取其中文字区域。对于文字，通过数学形态学的方法，获取逻辑独立的最大文本块，并记忆它们的位置属性信息，为阅读顺序确定做准备。

（3）文本块的阅读顺序确定

本文在前人版面理解研究的基础上，专注于文本区域阅读顺序确定的工作，基于日常书报杂志在版面结构上的普适原则，采用纵横切割法进行递归排序，直至不能分割为止；在利用普适原则之前，利用直线、方框信息将文本区域分成区块辅助排序；另外在遇到特殊结构时，如 L 形结构，不能使用纵横切割法进行排序，则建立相应的规则来进行阅读顺序确定。

在以上研究的基础上，本文提出了一种自底向上的基于数学形态学的分层自适应版面分析方法和一种分层的自顶向下的递归阅读顺序确定算法，通过提取和利用文档的版面信息，解析文档版面要素和排版规则，提升了读书机器人阅读日常书报杂志的能力。实验表明，本文的版面分析方法正确率为 96.4%，阅读顺序确定方法正确率为 96%，平均运算时

间为 0.162s，从而保证了读书机器人稳定，流畅地阅读日常书报杂志。

关键词：读书机器人，版面分析，阅读顺序确定，数学形态学，自适应，
分层

Document Layout Analysis and Reading Order Determination for a Reading Robot

ABSTRACT

The reading robot JoyTon makes use of optical, mechanical and electronic integration technology, converts paper documents to digital documents, and then converts them into sound information by TTS-technique. It is a robot made to help the visually-impaired people read. Optical character recognition (OCR) is usually used to transform paper documents to digital texts. Unfortunately, few document images obtained by a scanner or a digital camera allow for direct OCR processing. A noticeable reason for preprocessing before OCR is that the document images often include pictures, tables or some kind of special marks, or are structured in a complex layout. These elements and rules can not be understood by simple optical character recognition.

Layout analysis and understanding is the automatic recognition, segmentation and understanding process of image elements and their logical relationship. This paper will focus on the layout analysis and reading order determination algorithm of document images, including:

(1) Document Image Preprocess

The reading robot JoyTon gets document images by directly taking pictures of documents in the page-turning machine. Since the environmental factors are unstable, the images from the JoyTon robot suffer from ambient illumination, contrast stretch and shading effects, which will make the layout analysis much more difficult. So a series of image preprocessing is needed to extract single-page document images, improve image quality, dewarp geometric distortion, and carry out the image binarization and reduction.

(2) Layout Analysis

Firstly, the reduced binary image is scanned, and all connected domains are extracted, then lines, charts and texts can be distinguished according to the characteristics of domain size. In the second step they are processed in different ways. The lines will be directly used in reading order determination. The charts will be divided into tables, frame-texts and pictures according to their projection histogram charts. And mathematical morphology operations will be processed in the text domain to get maximum homogeneous text regions, whose location and property information will be kept to help determine the reading order.

(3) Reading Order Determination

This paper focuses on the reading order of the text part, and proposes a set of simple rules that cover the similarities of daily newspapers, magazines and books to build a recursive hierarchy algorithm. In this hierarchy algorithm, region separators have top priority, which means a region separator is prior to other information whenever they are detected at the same time. And certain layout configurations, such as L-shape, have lowest priority. And row pitch and column pitch detected by XY-cut are between them, while column pitch is prior to row pitch in light of western and Chinese composition convention.

Based on these researches, a bottom-up, morphology based, parameter-independent, hierarchy layout analysis and top-down recursive hierarchy reading order determination algorithm is proposed. JoyTon's ability to read daily newspapers, magazines and books is improved by extracting and using the document layout information. Experiments show that the correct ratio of layout analysis algorithm is 96.4%, the correct ratio of reading order determination method is 96%, and the average time consumption is 0.162s. The proposed algorithm will help to ensure that the reading robot reads daily newspapers, magazines and books stably and smoothly.

Keywords: a reading robot, layout analysis, reading order determination, morphology based, adaptive, hierarchy

目 录

用于读书机器人的版面分析和阅读顺序确定算法研究 1

用于读书机器人的版面分析和阅读顺序确定算法研究 I

摘 要..... I

ABSTRACT IV

第一章 绪论..... 8

 1.1 引言 8

 1.2 国内外研究现状 10

 1.3 本文研究内容 11

 1.4 本课题研究意义及应用前景 13

第二章 文档图像的预处理算法 14

 2.1 文档图像预处理算法的目的和流程 14

 2.2 初步版面分析 15

 2.3 图像增强和二值化 19

 2.3.1 图像增强..... 19

 2.3.2 图像二值化..... 21

 2.4 图像扭曲校正 25

 2.5 图像缩小和微膨胀 28

第三章 文档图像的版面分析 30

 3.1 版面分析的目的 30

 3.1.1 版面的几何结构分类..... 31

 3.1.2 本文版面分析的对象..... 32

 3.2 版面分析的研究现状 32

 3.2.1 版面分析的基本原理..... 32

 3.2.2 常用的版面分析方法..... 33

 3.3 分层版面分析的算法流程 33

 3.4 版面基本结构分类 34

 3.5 版面结构细分 40

 3.5.1 图表区域的分析..... 40

3.5.2 文字区域的分析..... 43

3.6 版面分析的成果 50

第四章 文本块的阅读顺序确定 52

4.1 阅读顺序确定的目的 52

4.2 阅读顺序确定的研究现状 53

4.3 阅读顺序确定的算法流程 54

4.4 阅读顺序确定算法 55

4.4.1 排版规则..... 55

4.4.2 基于树结构的阅读顺序确定..... 57

4.5 文本块的阅读顺序确定 63

5、实验结果与分析..... 66

5.1 对某一张报纸文档的处理结果 66

5.2 与其他算法的比较 70

6、总结与展望..... 72

6.1 本文主要研究工作及成果 72

6.2 结论与展望..... 73

参 考 文 献..... 76

致 谢..... 79

攻读硕士学位期间已发表或录用的论文 81

第一章 绪论

1.1 引言

世界卫生组织估计，目前全世界约有一亿八千万人患有视觉残疾，其中约四千五百万人完全失明。如不进行有效的干预，盲人数量在 2020 年将增至 7500 万。在当今信息繁荣，社会高速发展的时代，盲人和视力障碍者同样需要享受信息财富，进行学习、娱乐和工作。然而，享受信息财富的第一步，阅读，尤其是阅读纸质出版物，对于他们却充满了层层障碍。



图 1-1 读书机器人 JoyTon

Fig.1-1 The Reading Robot JoyTon

近年来，作者所在的实验室在助盲方面做了一些基础性的研发工作，先后开发了自动翻页机器人^[1]、翻页扫描机器人^[2]和读书机器人^[3]，致力于帮助盲人和视力障碍者克服阅读的障碍，丰富他们的文化生活。作者所在实验室自主研发的读书机器人 JoyTon，如图 1-1 所示，是助盲机器人中的佼佼者^[4, 5]。它能够不借助人的干预，自

动地通过光机电一体化系统，将纸介质信息数字化，然后通过文本语音转换，将视觉信息转换为听觉信息，为盲人和视力障碍者提供人性化的阅读服务。JoyTon 的工作流程如图 1-2 所示。

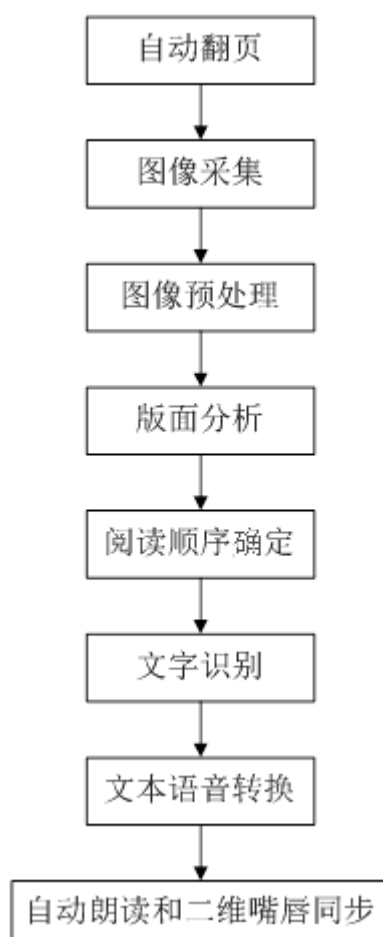


图 1-2 JoyTon 的工作流程图

Fig.1-2 The Workflow of JoyTon

在将纸介质信息数字化的过程中，使用到的一项关键技术是光学字符识别。光学字符识别（又称 Optical Character Recognition, OCR）^[6-8]，经过几十年的发展，对单个字符的识别技术已经比较成熟，可以迅速有效地获取文本。然而，除了文字以外，文档中的版面格式也是信息的重要组成部分，仅依靠光学字符识别技术，读书机器人

只能正确地阅读极少数的文档。究其原因，其中很重要的一方面是，文档图像通常会包括图像、表格、文字、线框等要素，并以各种版面规则进行排列，这些要素和规则，是简单的光学字符识别所不能理解的。文档图像的版面分析和阅读顺序确定是对图像内的文本、图像、表格、线框等要素的特征和位置关系进行自动的识别、分割和理解的过程，它按照文档的逻辑顺序为读书机器人提供逻辑独立的纯文本图像区域，是读书机器人进行自主文档阅读的重要步骤，其结果影响到整个读书机器人系统的效能。

因此，本文将着重对文档图像的版面分析和阅读顺序确定算法进行研究，通过提取和利用文档的版面信息，解析文档的要素和规则，提升读书机器人阅读日常书报杂志的能力。

1.2 国内外研究现状

版面分析和阅读顺序确定，可以看作一个特定的图像理解问题，其研究内容涉及图像处理、模式识别、情报检索和人工智能等多学科的交叉领域。版面分析通过对要素特征的理解和对版面几何结构的分析，从文档图像中提取逻辑独立区域的位置信息，并把它们区分为文本、图像、表格、线框；阅读顺序确定，是版面理解问题之一，通过对线框信息和位置关系的分析，解读版面规则中蕴含的逻辑关系，对逻辑独立的纯文本区域进行排序。

版面分析与版面理解的技术已有几十年的历史^[9, 10]。八十年代初期，版面分析与理解的研究主要是对布局比较简单的文本版面进行分割；到了九十年代，版面理解的研究才开始逐步展开^[11]；目前，版面分析与理解技术的研究，已经可以对不同领域的版面，采用不同的技术进行研究，并逐步得到实际应用。国内外许多研究机构，都在版面分析与理解领域建立了研究中心，并且取得了一定的研究成果^[12]，这些研究机构从不同角度，对版面分析与理解技术的算法进行了研究，其中比较著名的有：

德国 DFKI 研究中心，共有五个主要研究小组，其中之一是由著名的版面分析专家 Dr. A. Dengel 领导的信息管理与文档分析小组。

加拿大 Concordia 大学模式识别与机器智能研究中心，在国际字符识别权威 Dr. C.

Y. Suen 的带领下，在字符识别、版面分析与理解以及自然语言理解等方面的研究居于世界领先水平。

美国 Nevada Las Vegas 大学的信息科学研究所，主要任务是提高版面理解的自动化技术。

美国 Washington 大学的智能系统实验室，最早开发了用于 OCR 和版面分析与理解算法的 CD-ROM 版面数据库，该数据库包括了各文本、图像版面的图像，以及相应对象，如字符、符号和线条等的有关信息，为系统、客观的评价版面分析与理解过程中各个算法提供了方便。

国内起步较晚，到九十年代才逐步开始了对版面分析和理解的研究^[13]。由于中文版面设计比较自由，版面内文本、图表交错，各个文本区域字体、字号频繁变化，间隔线、花边线种类繁多，导致中文版面分析需要解决更多的困难，随着中文出版物形式越来越多，版面千变万化，版面分析与理解技术已成为 OCR 系统中的关键问题。

目前，中文版面分析与理解的研究已经取得了一批可供实际应用的研究成果，例如，中科院自动化所的邮政编码识别机，北京邮电大学的银行票据图像处理与识别引擎，汉王的表格自动录入系统，清华紫光的 OCR 系统，良田公司的良田高拍仪等。这些研究成果在汉字识别速度和效率上，基本上能够达到令人满意的效果，但是在版面分析与理解方面，不论是分析速度还是准确性都有待提高。为了提高版面分析与理解的正确率，往往需要用户根据对象从一些常用版面格式中选择，甚至由用户标示区域和属性，这些人工干预的方法会降低 OCR 系统的自动化程度，使整个系统在实际应用中受到限制。

综上所述，目前版面分析与版面理解的研究已经广泛展开，并且提出了许多有价值的概念，在某些细分领域，比如表格、信件等的分析和理解，取得了有实际应用价值的成果，也为本文的研究提供了理论上和技术上的基础。

1.3 本文研究内容

本文的目标是研发一种具有实际应用价值的版面分析和阅读顺序确定算法，使读

书机器人 JoyTon 能够正确流畅地阅读大部分日常书报杂志。因此，本文的研究除了需要保证版面分析和阅读顺序确定算法的正确性之外，还必须解决如下几个问题：

1、抗干扰性

由于读书机器人获取文档图像时所处的环境存在各种可能，文档图像往往受到光照不均匀，对比度拉伸，物体阴影的影响，较之扫描获得的图像存在更多的噪声；同时，文档版面同样可能存在污点，图像可能倾斜扭曲。这就要求本文的版面分析和阅读顺序确定算法必须具备较高的抗干扰性。

2、保持整体性

版面分析过程中，对文本块的划分，需要在尽可能将逻辑关系一致的文本区域合并，即取得逻辑独立的最大文本块，以保持版面分布的整体特征，减少阅读顺序确定和光学字符识别的运算量。

3、高适应性

随着出版物的种类增多，即使是日常的书报杂志，版面设计也达到了相当的复杂程度，复杂度不仅体现在版面几何结构上，而且体现在文本的字体字号上。为了使读书机器人能够自由地在不同阅读物间切换，且不需要人工干预，读书机器人应当尽量减小对经验阈值的依赖性，采用根据当前版面图像的统计信息自动获取参数的方式，提高算法的适应性

4、高效性

要想在读书机器人上取得实际应用价值，版面分析和阅读顺序确定算法必须是快速、高效的，从而尽量减少翻页后的停顿时间，使读书机器人能够流畅地阅读。因此，本文要尽量提高处理速度，降低版面分析和阅读顺序确定算法的复杂性。

为了解决如上问题，本文将首先对文档图像进行预处理，然后展开对版面分析和阅读顺序确定算法的研究，本文结构如下：

1、第二章中将介绍文档图像预处理算法的流程，并对初步版面分析，图像二值化，图像扭曲校正，以及图像缩小和微膨胀进行阐述。

2、第三章中将详细介绍文档图像的版面分析算法。在对版面分析的研究现状进

行回顾之后,将先给出本文分层版面分析方法的整体视图,然后详细描述这种分层的,基于数学形态学的,自底向上的,自适应的算法。

3、第四章中将详细介绍文本块的阅读顺序确定算法。在对阅读顺序确定算法的研究现状进行回顾之后,将先给出本文分层阅读顺序确定算法的整体视图,然后详细介绍这种分层的,基于纵横切割法的,自顶向下的递归算法。

第三、四两章是本文的重点。

4、第五、六两章中将对实验结果进行分析,并给出总结和展望。

1.4 本课题研究意义及应用前景

本文对文档图像的版面分析和阅读顺序确定算法的研究,通过提取和利用文档的版面信息,解析文档的要素和规则,将提升读书机器人阅读日常书报杂志的能力。对于全世界患有视觉残疾的一亿八千万人,本文的研究,将为他们通过阅读,进行学习、娱乐和工作,享受信息财富提供新的更加稳定亲和的可能性。

本文提出的版面分析和阅读顺序确定算法,不仅能够用于读书机器人,为盲人服务,而且能够提高纸质出版物数字化过程中的自动化程度与工作效率。

第二章 文档图像的预处理算法

2.1 文档图像预处理算法的目的和流程

由于读书机器人 JoyTon 的文档获取机制，是控制数码相机对翻书机构中的纸质文档直接拍照，导致获取图像时的环境因素多变，文档图像往往包含背景，并且受到光照不均匀、对比度拉伸、物体阴影的影响，为版面分析和阅读顺序确定增加了难度，所以需要进行一系列的图像预处理，从而提取单页文档图像，提升图像质量，进行几何扭曲校正，并且完成图像二值化和缩小。这些预处理操作，将会提升版面分析和阅读顺序确定算法的抗干扰能力，并且为提升处理速度保证读书机器人流畅工作做好准备。

本文的图像预处理算法流程如图 2-1 所示。首先对读书机器人 JoyTon 获取的原始页面图像进行初步的版面分析，去除背景，获得单页的文档图像；然后进行图像增强，提升图像的质量；接下来进行图像二值化，去除文档图像中的多余信息；再然后进行图像几何校正，缓解图像的倾斜和扭曲，从而提升版面分析和光学字符识别的正确率；最后，对无变形图像的二值结果进行缩小，从而减少版面分析过程的计算量。

本章接下来的部分，将对初步版面分析，图像增强和二值化，图像扭曲校正，以及图像缩小和微膨胀进行阐述。

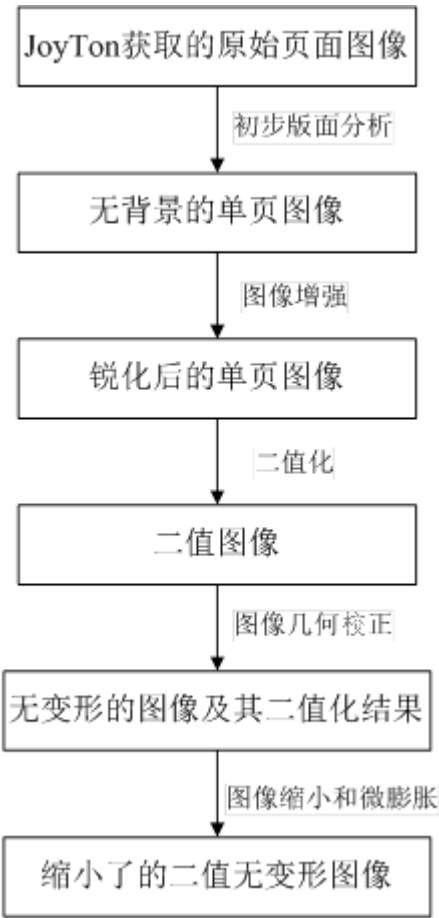


图 2-1 文档图像预处理算法流程图

Fig.2-1 Workflow of Document Image Preprocessing Algorithms

2.2 初步版面分析

由于读书机器人 JoyTon 是通过控制数码相机对翻书机构中的纸质文档直接拍照的，所以通常获得的文档图像是包含托盘等背景的双页图像，如图 2-2 所示。初步版面分析的目的是去除背景，并且提取两个单页图像。

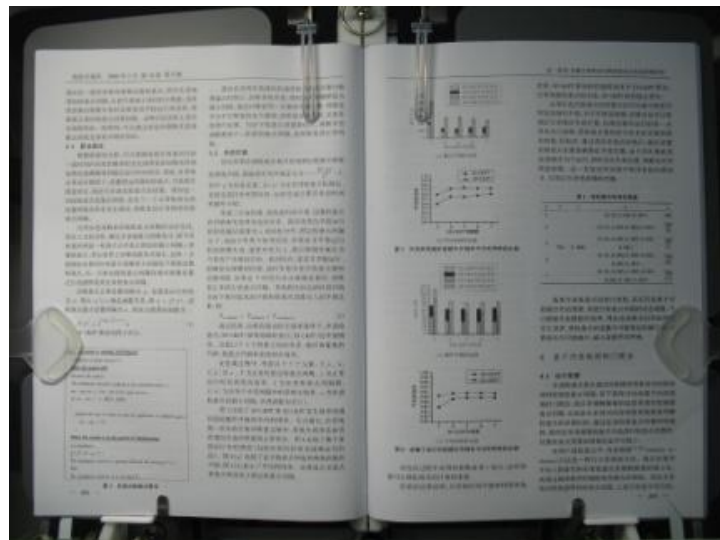


图 2-2 初始文字图像

Fig.2-2 Source Page Image from the Camera

Nagy、Seth 和 Viswanathan 在 1992 年提出了一个经典的版面分析的算法——纵横切割法^[14]，它以文字图像在水平方向和垂直方向的投影函数为基础，通过分析函数曲线的走向，来判断文档的结构信息。本文中的初步版面分析借鉴了这一思想，具体的流程图如图 2-3 所示。

由于数码相机获得的原始图像像素比较大，为 2880*1720，如果直接对原图进行操作，将会花费不菲的时间；而从图 2-2 中可以发现，原始图像的背景色为深色，页面为浅色，区分相对比较明显，对原图进行适当的缩小，可以在保证初步版面分析结果稳定的情况下，最大幅度的提高处理速度。因此按照公式 2-1 进行图像缩小：

$$\begin{aligned} x &= k * x_0 \\ y &= k * y_0 \\ f(x, y) &= g(x_0, y_0) \end{aligned} \quad (2-1)$$

其中，(x, y)代表缩放后的图像坐标，(x₀, y₀)代表对应位置的原图坐标，f、g 代表像素值，k 为缩放比例。在这里，k 值是完全由翻书机构的尺寸和数码相机的分辨率决定的，经过反复试验，对于本文的环境，0.05 是最佳取值。

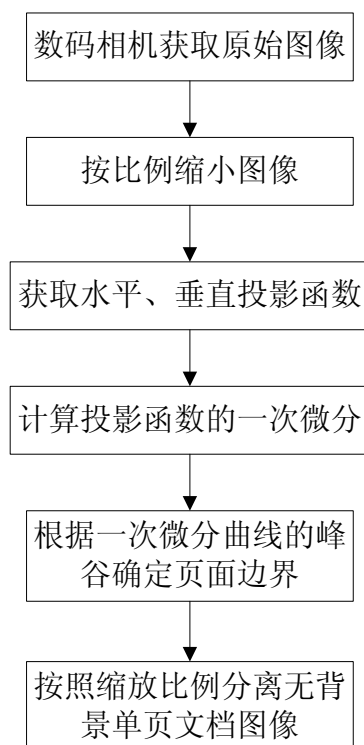


图 2-3 版面分析流程图

Fig.2-3 Workflow of Preliminary Document Analysis

由于背景为单一的深色托板,而目标物书本一般为白底或浅色底,在其交界处,垂直和水平方向的投影直方图将会表现出明显的突变特征,因此得到图像水平和垂直投影函数的一阶微分曲线后,就可以根据其波峰和波谷找到页面的边界。缩放后图像水平和垂直投影函数的一阶微分曲线如图 2-4 所示,水平投影一阶导数曲线 (a) 上的一个波峰和一个波谷分别代表了页面的上边缘和下边缘,垂直投影一阶导数曲线 (b) 上的两个波峰和一个波谷分别代表了页面的左边缘,中缝和右边缘。



图 2-4 图像水平和垂直投影函数

(a) 水平投影函数一阶微分曲线示意图 (b) 垂直投影函数一阶微分曲线示意图

Fig.2-4 Vertical and Horizontal Projection Histogram of the image

(a) First Derivative of Horizontal Projection Histogram

(b) First Derivative of Vertical Projection Histogram

在按照缩放比例分离无背景单页文档图像时,应当特别注意保证单页文档图像的像素数是 8 的倍数,以避免在后续的处理过程中程序运算出现不必要的错误。图 2-5 是最终分割得到的两幅页面图像。

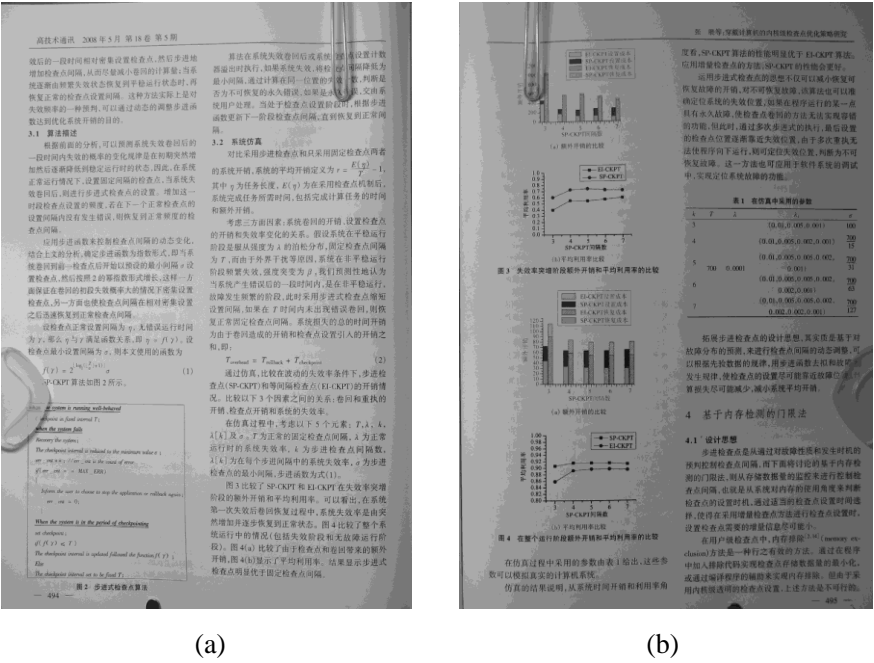


图 2-5 最终分割得到的页面图像

(a) 分割后的左半页 (b) 分割后的右半页

Fig.2-5 The Segmentation Result: Page Image

(a) The Left Half Page Image

(b) The Right Half Page Image

2.3 图像增强和二值化

对于读书机器人 JoyTon 来说，一幅文档图像，有价值的信息集中在文字和图表的部分，其他的信息反而会对 JoyTon 的正常阅读造成干扰，降低其效率或者正确率。本文中采用图像增强，来提升文档图像的质量；采用图像二值化，来去除文档图像中的多余信息。这两个操作有一个共同的目的，就是将文档图像中的文字和图表加以强化和提取。

2.3.1 图像增强

图像增强是数字图像处理的基本内容。在本文中，将采用图像增强技术中的图像锐化，增强文字和背景的对比度，突出文档图像中的文字和图表，削弱或除去背景噪声，使得到的新的图像比原始图像更适合机器人 JoyTon 的阅读。

在去噪声处理过程中，不可避免的会使文档图像的文字和图表边缘发生模糊。采用图像锐化处理就能有效的加以改善。若把平滑理解为积分作用，则图像锐化就相当于微分作用。图像锐化中使用最广泛的是拉普拉斯微分算子^[15]，它是一种具有各向同性(即旋转不变性)的线性算子。

拉普拉斯微分算子的定义及其差分形式如下：

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (2-2)$$

$$\nabla^2 f = \Delta_x^2 f + \Delta_y^2 f \quad (2-3)$$

对于离散数字图像 $f(x, y)$ ，由文字扩散现象引起的图像模糊，可以用下式进行锐化：

$$\begin{aligned}
\nabla^2 f &= \Delta_x^2 f + \Delta_y^2 f = [\Delta_x f(x+1, y) - \Delta_x f(x-1, y)] + [\Delta_y f(x, y+1) - \Delta_y f(x, y-1)] \\
&= [f(x+1, y) + f(x-1, y) - 2f(x, y)] + [f(x, y+1) + f(x, y-1) - 2f(x, y)] \\
&= f(x+1, y) + f(x-1, y) + f(x, y+1) + f(x, y-1) - 4f(x, y)
\end{aligned} \tag{2-4}$$

$$g(x, y) = f(x, y) - k\nabla^2 f(x, y) \tag{2-5}$$

这里 k 是与扩散效应有关的重要系数。若 k 取值过大，则图像轮廓边缘会产生过冲现象；若 k 取值过小，则锐化效果不明显。

处理中，我们使用了公式（2-6）中第二个模板矩阵 T 对每一像素点的 R 、 G 、 B 三色分量同时进行修正，其修正的算法模型为如下的卷积公式，具体的算法流程如图 2-7 所示。

$$T = \begin{bmatrix} -1 & -1 & -1 \\ -1 & 9 & -1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & -2 & 1 \\ -2 & 5 & -2 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{2-6}$$

图 2-6 为一幅图像局部的锐化前后对比，可以看到，图像经锐化变换后，文字笔画更加清晰、线条更加连续，文档的整体效果明显增强。

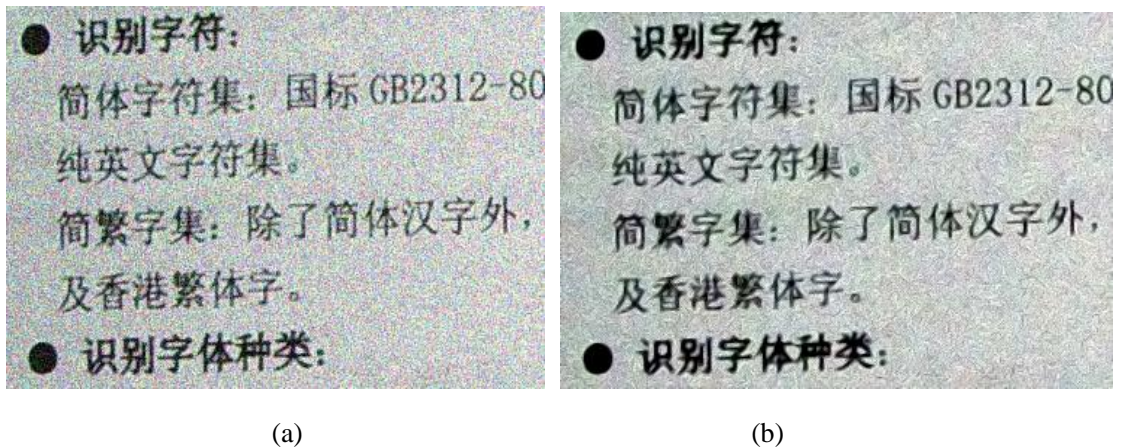


图 2-6 图像增强技术效果示意图

(a) 局部原始图像部分 (b) 锐化处理结果

Fig.2-6 Image Enhancement

(a) Part of an Original Image (b) The result of Image Sharpen

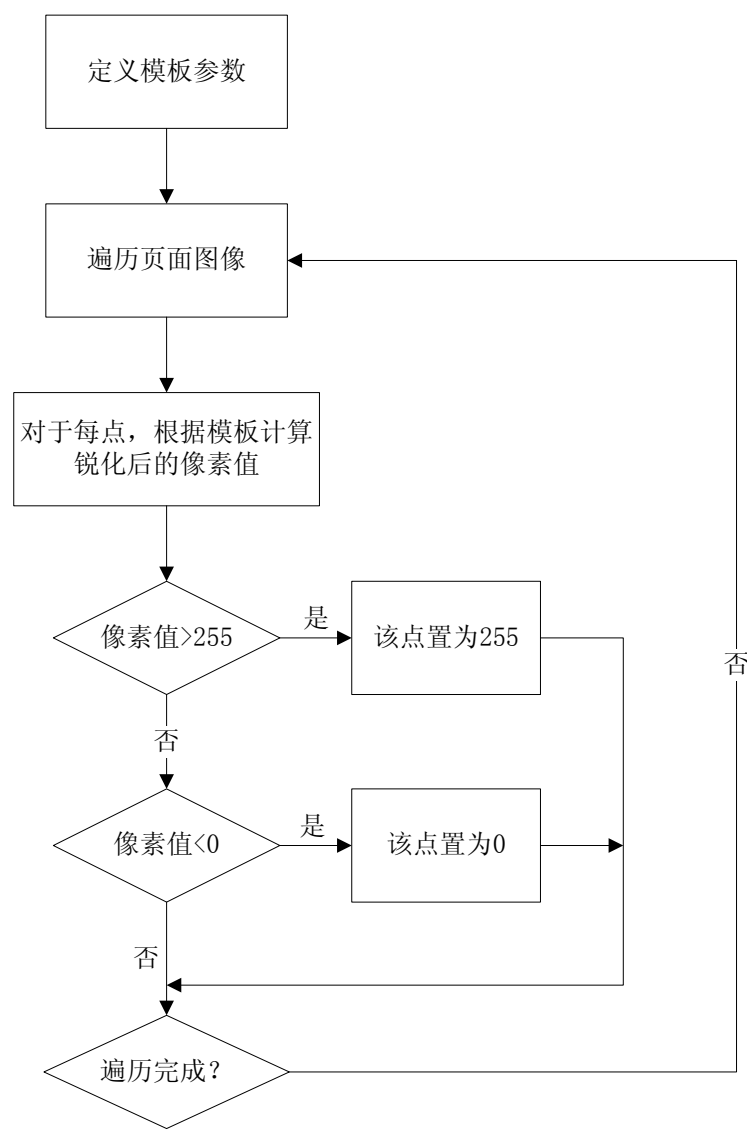


图 2-7 图像锐化算法流程

Fig.2-7 Workflow of Image Sharpen Algorithm

2.3.2 图像二值化

由于二值图像只有两个灰度等级，但却足够描述文字的纹理，所以图像二值化运算，将能够在完整保存文字纹理，图表位置属性信息的基础上，有效地降低后续运算

的复杂度。因此二值化不仅是光学字符识别的前提，也是本文版面分析与阅读顺序确定算法的基础。

一般而言，图像二值化方法可以分为基于全局的阈值选取算法和基于局部的阈值选取算法^[16]。由于读书机器人 JoyTon 获取图像时的环境因素多变，文档图像往往受到光照不均匀、对比度拉伸、物体阴影的影响，所以本文将采用虽然需要更大的运算量，但是对环境适应能力更强的局部阈值选取算法。

最常见的基于局部的阈值方法是 NiBlack 算法^[17]，NiBlack 算法强调二值化门限值随着区域的变化而动态调整，利用一个模板窗口不断去寻找局部的最佳阈值。在光线环境不确定的情况下，会获得比较好的结果。像素点(x, y)的局部阈值选择公式如下所示：

(2-7)

其中， $m(x, y)$ 代表模板窗口内的平均灰度值， $s(x, y)$ 代表模板窗口内的方差，参数 k 和模板窗口的大小都是经验参数，本文中，采用的经验值为：窗口大小 $n=7 \times 7$ ，参数 $k=0.5$ 。

但是，实际运用中，经常会出现整体光线环境比较暗或存在阴影的情况，此时，NiBlack 方法所获得的结果不会很理想。因此我们参考 Sauvola 的方法^[18]，对 NiBlack 方法进行如下的改进：

(2-8)

其中 R 为归一化参数。窗口大小和 k 保持不变。

图 2-8 是具体的二值化算法流程。

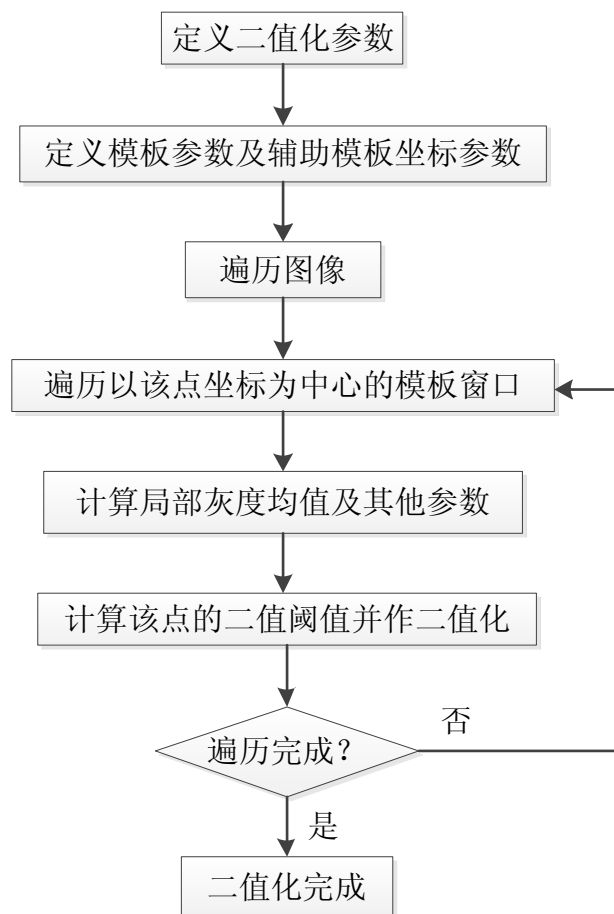


图 2-8 二值化算法流程图

Fig.2-8 Workflow of the Binarization Algorithm

图 2-9 中是对读书机器人中实际抓取的图像进行多种二值化运算的结果。(a)为全局化的 OSTU 方法得到的结果，(b)为本文采用的 Sauvola 方法得到的结果。在背景色渐变的情况下，本文采用的基于局域的方法效果很好。(c) 是从原图像上光线较暗的一部分，(d)、(e) 分别为 Niblack 方法和本文采用的 Sauvola 方法的二值化结果，在这样的光线条件下，本文采用的方法比 NiBlack 方法适应力更强。

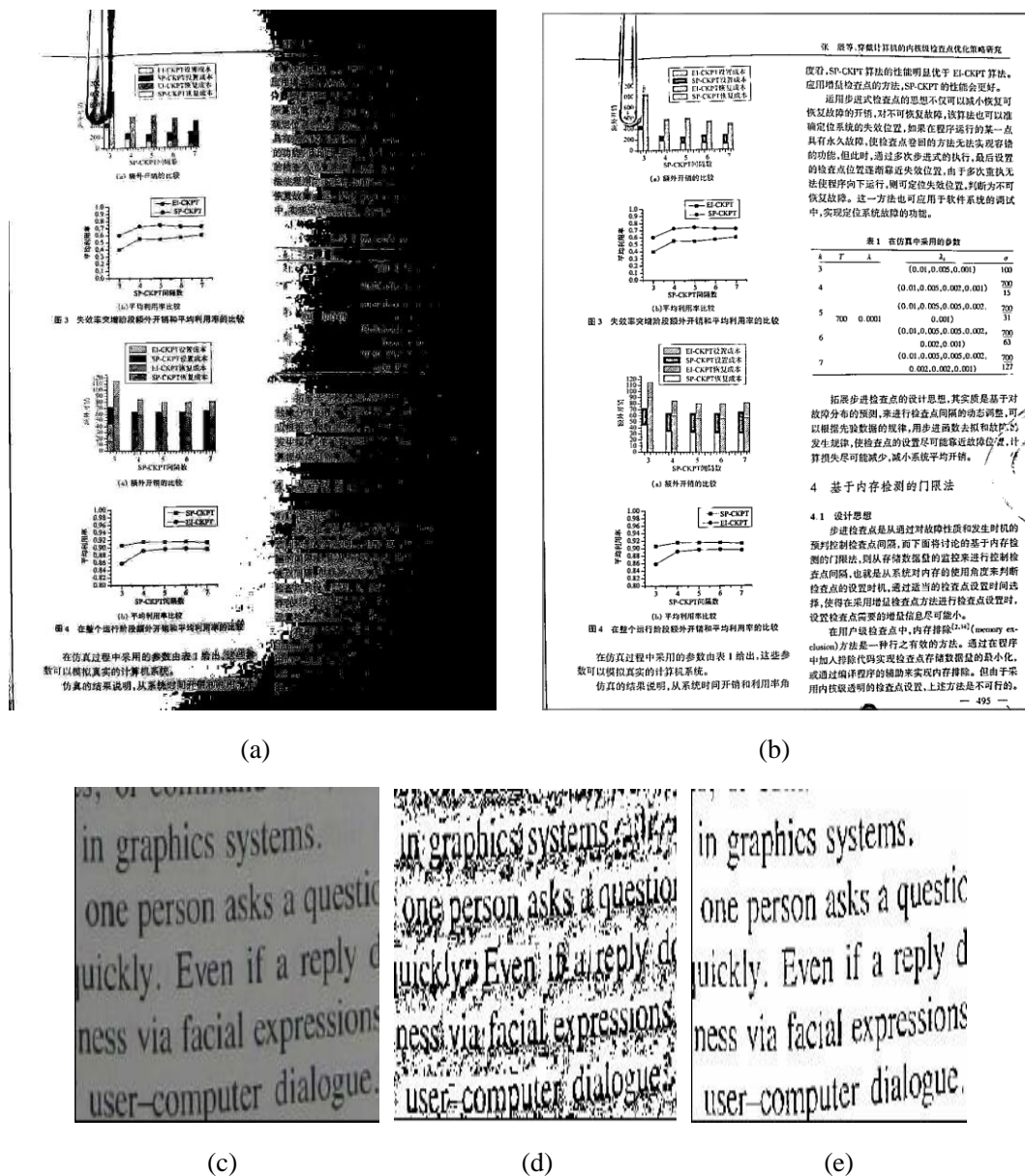


图 2-9 二值化运算结果

(a) OSTU 方法的结果 (b) 本文 Sauvola 方法的结果

(c) 光线较暗的部分图像 (d) NiBlack 方法的结果 (e) 本文 Sauvola 方法的结果

Fig.2-9 Binary Results

(a) Result of OSTU Method (b) Result of Sauvola Method

(c) The Darker Part of a Document Image (d) Result of NiBlack Method

(e) Result of Sauvola Method

2.4 图像扭曲校正

读书机器人获取的文字图像,由于书的厚度和机器人视觉系统的视角限制,或多或少的存在着一些几何形变,这些畸变在页面的边缘显得尤为明显,如图 2-10 所示。几何形变将带来的问题有两个,版面分析出错和光学字符识别率下降,因此需要进行图像扭曲校正,使版面和文字恢复到没有形变的状态。

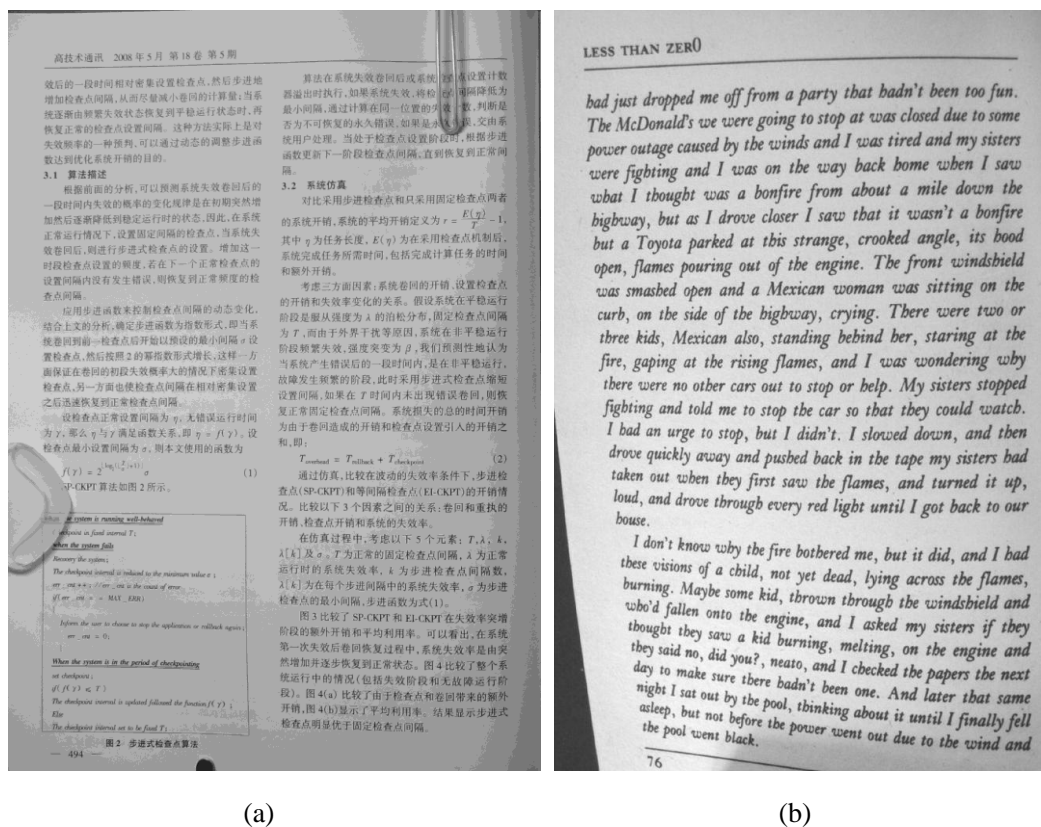


图 2-10 文字图像中的几何形变

(a) 普通扭曲 (b) 严重扭曲

Fig.2-10 Distortion in Page Image

(a) Normal Distortion (b) Serious Distortion

本文作者所在的实验室，在图像几何扭曲校正方面已经进行了大量的研究，作者的学长张伟业硕士结合前人的工作和实际的实验环境，提出了如图 2-11 的扭曲校正流程^[19]。该算法根据读书机器人中的页面实际情况，首先分析文字图像的连通域，定位文字行，然后用三次样条曲线来拟合文字行，最后根据顶部和底部的两条曲线作为基准来复原没有扭曲的图像。

几何扭曲校正的实验结果如图 2-12 所示

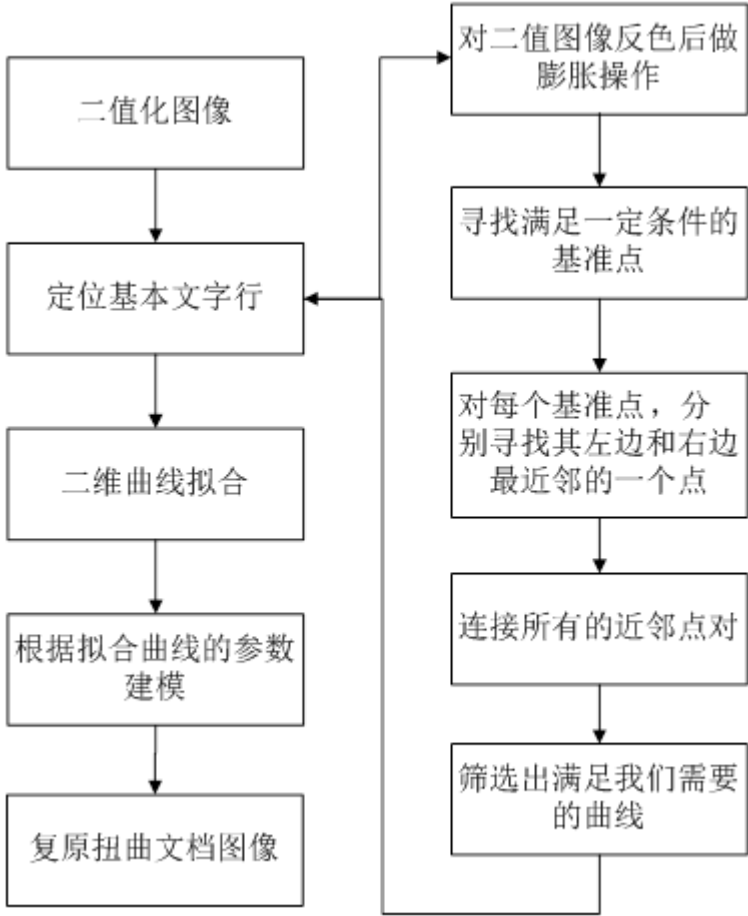


图 2-11 图像扭曲校正流程

Fig.2-11 Image dewarping workflow

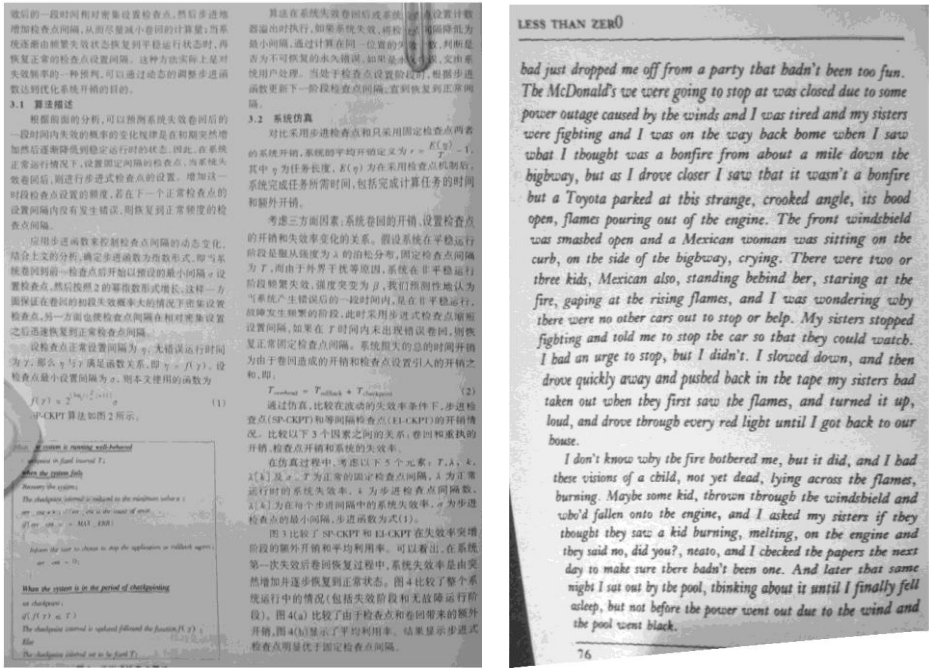


图 2-12 几何校正结果
Fig.2-12 Dewarping Results

值得注意的是，当原始页面图像分栏，如图 2-10(a)，或者包含了大量的图像表格信息时，本文中的图像扭曲算法在寻找文字行这一步骤时不一定能保证找到连续的贯穿整个篇幅的文字行，从而降低图像几何扭曲校正的效果。如果在图像扭曲校正之前，进行一定的版面分析，然后对各栏分别进行图像扭曲校正，可以得到更好的效果。

然而，当原始图像变形严重，版面结构复杂时，几何变形将模糊图像各栏的边界，如果不能在版面分析之前，进行足够的图像扭曲校正，版面分析的出错率也会大大上升。

综合以上两个方面，本文采用首先进行初步的版面分析，然后进行图像几何扭曲校正，最后进行精细版面分析和阅读顺序确定的基本流程。同时，由于图像几何校正的适应范围不够广泛，这里也给读书机器人提出了保证文档图像正位的要求。

2.5 图像缩小和微膨胀

由于版面分析并不需要文字、图表等要素的细节纹理，而图像缩小可以大量减少版面分析的运算量，所以在预处理阶段进行一步图像缩小的操作，并且采用适当的算法，使得图像缩小过程中，对文字等白色像素组成的图像前景进行少量的膨胀。

在本章第三节中提到，读书机器人 JoyTon 获取图像时的环境因素多变，文档图像往往受到光照不均匀、对比度拉伸、物体阴影的影响，本文为了保留完整的文字纹理，采用针对整体光线环境偏暗的二值化方法。这就导致一部分的前景像素（白色）被当作了背景像素（黑色）处理，原本应当连接的直线线条、文字比划等发生了断裂。所以同一般的隔点采样不同，我们采用了基于块结构的图像缩小。

缩小后图像的每一个像素代表原始图像中一个 $N \times N$ 像素的图像块。只有当对应的 $N \times N$ 个像素都是黑色(背景)时，如图 2-13(a)，缩小后图像的这个像素才为黑色(背景)，否则，如图 2-13(b-e)，为白色（前景）。这种图像缩小的方法，能够连接被少于 $2N-2$ 个像素点分隔的白色区域，不仅在减少了版面分析过程中目标运算像素数量，如果运用得当，还可以大量减少初始连通域的数量。

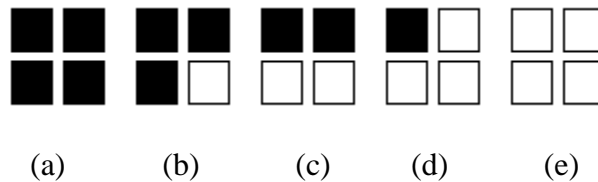


图 2-13 图像块示意图

Fig.2-13 Image Block

通常情况下，本文默认采用 2×2 的块对文档图像进行缩小处理。在下一章所述版面分析过程中，可以通过连通域高度的统计数据，来得到对应文档文字行的高度 H ，计算

$$N = \frac{H}{8} + 1 \quad (2-9)$$

如果 $N > 2$ ，则使用 $N \times N$ 的块重新进行这一步操作。

使用进行图像缩小的结果如图 2-14 所示，可以明显看到文字、图表的膨胀效果。

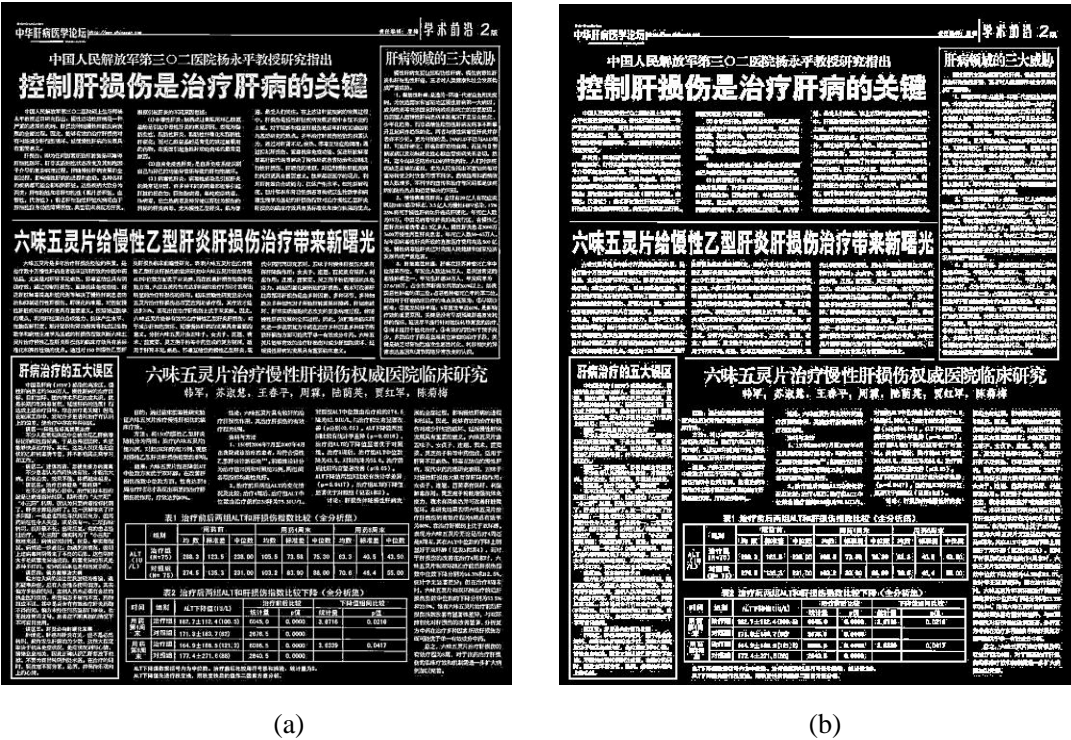


图 2-14 基于块结构的图像缩小示意图

(a) 缩小前图像 (b) 图像缩小结果

Fig.2-14 Image Reduction Based on Block Structure

(a) Image before Reduction (b) Reduction Result

第三章 文档图像的版面分析

3.1 版面分析的目的

读书机器人 JoyTon 的目标是流畅地阅读日常书报杂志，简单的如书籍、技术文档，复杂的如杂志、报纸都是读书机器人的阅读对象。为了达成这个目标，它需要有能力阅读各种日常版面结构的图像，一个报纸图像如图 3-1 所示。



图 3-1 一个报纸图像
Fig.3-1 A Newspaper Image

版面分析是纸介质信息数字化的首要步骤，其目的在于提高数字化处理的智能程度，通过对要素特征的理解和版面几何结构的分析，从文档图像中提取逻辑独立区域

的位置信息，并把它分为文本、图像、表格、线框，为阅读顺序确定提供各区域位置、属性等相关信息。

3.1.1 版面的几何结构分类

从几何结构出发，版面可分为矩形版面和非矩形版面两类。其中矩形版面可分为无嵌入式和嵌入式两种^[20]。

1. 矩形版面

a) 无嵌入式矩形版面

无嵌入式矩形版面的特点是，每个版面对象的外切矩形互不相交，即其外切矩形没有任何重叠或交叉，版面对象之间的几何结构为上下结构或左右结构等简单位置关系。

b) 嵌入式矩形版面

嵌入式矩形版面的特点是，版面对象的外切矩形有相交情况，各个版面对象在位置上存在包含关系，几何结构为内外结构，其最基本的结构类型为如图 3-2 所示的两种。

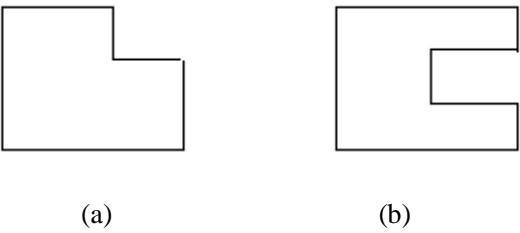


图 3-2 嵌入式矩形版面的基本类型

(a) 角嵌入式版面 (b) 边嵌入式版面

Fig.3-2 Basic Types of embedded Rectangular Layout

(a) Corner embedded Layout (b) Edge embedded Layout

这两类基本类型嵌套排列，可以生成复杂的嵌入式版面。

2. 非矩形版面

非矩形版面的特点是，版面对象不能用矩形描述，这类版面使用量少，而各个非矩形版面共性少变化多，情况复杂。这类版面通常需要专门的版面分析和理解方法来处理^[21, 22]。

3.1.2 本文版面分析的对象

由于日常书报杂志基本都采用矩形版面，本文的版面分析，也将针对矩形版面展开，而对于个性化更强，分析过程中需要付出大量额外运算的非矩形版面，在本文将不予考虑。

3.2 版面分析的研究现状

3.2.1 版面分析的基本原理

版面分析的技术已有几十年的历史，其研究已经可以对不同领域的版面，采用不同的技术进行研究，发表了大量的学术论文，并逐步得到实际应用。究其根本，版面分析算法可以按照分析过程中采用的特征，可以分为基于形状、基于纹理和基于神经网络三类。

基于形状的方法^[23]，是利用文本、图像、表格、线框等版面要素，具有不同的形状规律进行版面分析的方法。这种方法的优点是，采用的特征容易获取，逻辑简单，且对于不同的文档具有较高的适应能力；缺点是，不能够正确识别形状特殊的版面要素，且对于背景噪声的适应能力较差。

基于纹理的方法^[24]，通过小波运算来提取纹理信息，再根据文字和线条具有特殊的纹理特征，来提取和区分文本、图像、表格、线框等版面要素。这种方法的优点是，能够适应复杂的背景环境。缺点是，运算量大，耗时长^[25]。由于其抗噪能力强的特点，这种方法经常被应用于海报的分析。

基于神经网络的方法^[26]，通过样本图像对神经网络进行训练，从而使神经网络具有识别文本、图像、表格、线框的功能。这种方法的优点是，对于属性相对一致的文档图像，能够进行较好的分析。

3.2.2 常用的版面分析方法

版面分析的算法，按照逻辑独立区域获取方式的不同，可以分为自顶向下的方法、自底向上的方法和混合型方法。

自顶向下的方法^[27]，从版面整体出发，通过纵横投影的方法，搜索各个区域间的空白条带，并把这些空白条带作为各个区域的边界。这种方法的优点是运算简单，实现迅速；但是需要有贯穿文档图像的空白条带，且空白条带要符合一定的规律排布。

自底向上的方法^[28, 29]，先获得最底层的版面信息，进行数学形态学的变换和连通域的搜索，逐步合并相同属性的连通域，从而得到逻辑独立的区域，然后根据要素的特征进行区分。这种方法的优点是，能够处理复杂的版面结构；但是运算复杂，速度较慢。

混合型的方法^[30]，将自顶向下和自底向上的版面分割方法结合起来，同时使用了全局和局部信息来处理比较复杂的版面，以充分发挥两者的优势，获得相对优越的性能。

此外，利用图像处理技术对版面进行分割的方法，例如采用基于特征矢量的纹理分析、利用小波变换和神经网络技术等，只需一步就可将版面分成所需的块，可处理具有复杂几何结构的版面。但是该类方法共同的缺点是计算量非常大而准确度又不够高

3.3 分层版面分析的算法流程

根据本文第一章第三节的分析，读书机器人 JoyTon 中应用的版面分析方法，需要具备这样几个特征：准确性，抗干扰性，保持整体性，高适应性，以及高效性。本

文根据读书机器人的特性和前人的研究，提出了一种分层的，基于数学形态学的，自底向上的，自适应的算法。其流程如图 3-3 所示。

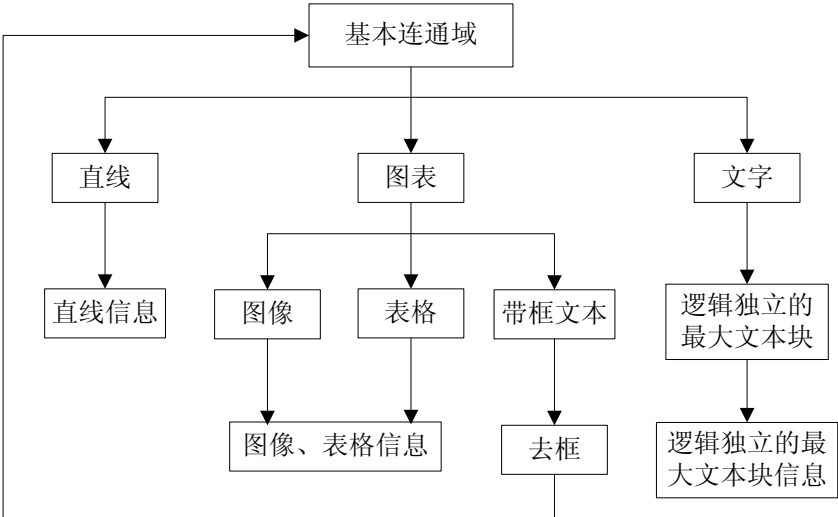


图 3-3 分层版面分析的算法流程图

Fig.3-3 The Workflow of Hierarchy Layout Analysis Algorithms

这种方法，首先扫描预处理获得的二值缩小图像，抽取文档图像上所有的连通域，然后根据连通域的尺寸特征，区分直线、图表和文字。对于直线，留作阅读顺序确定的信息。对于图表，根据它们的纵横投影，可以区分为图像、表格和带框文本；带框文本在去框后，作为单独图像进行分析，获取其中文字区域；对于文字，通过数学形态学的方法，获取逻辑独立的最大文本块，并记忆它们的位置属性信息，为阅读顺序确定做准备。

3.4 版面基本结构分类

本节中，首先采用区域生长的算法抽取二值缩小图像中所有的连通域，绘制连通域边界盒，并对重叠的连通域进行合并，然后根据连通域的尺寸特征，将它们区分成直线、图表和文字，为版面结构的细分做好准备。基本流程如图 3-4 所示。

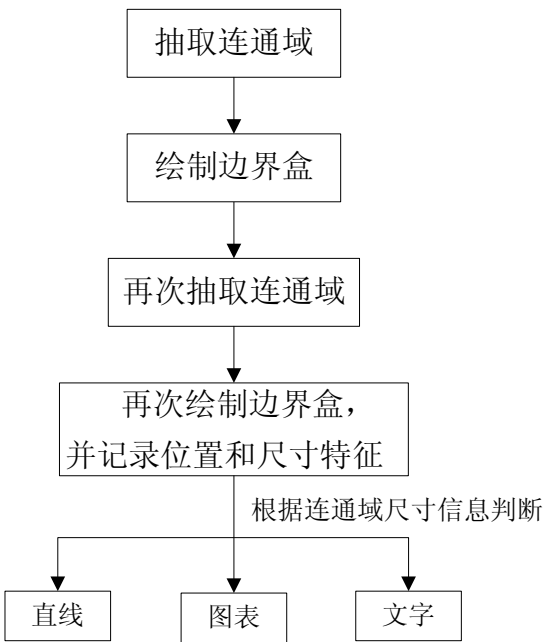


图 3-4 版面基本结构分类流程图

Fig.3-4 Workflow of Basic Layout Element Classification

3.4.1 连通域的抽取和边界盒的绘制

扫描二值缩小图像，抽取连通域，可以使用区域生长的算法来进行。区域生长是一种根据事先定义的准则，将像素或子区域聚合成更大区域的过程。基本方法是，以一组种子像素点开始，将与种子性质相似的相邻像素附加到生长区域的每个种子上。

在本文中首先确定图像最左上角的前景像素（白色）位置，然后以它为种子进行生长；在二值图像中，相似规则非常简单，就是与种子等值。当没有像素满足相邻且与种子等值的条件时，一个连通域的提取就完成了。然后按照同样的规则抽取图像中所有的连通域。

连通域是进行版面分析的基本单元，连通域的边界盒是连通域的重要特征，这个特征足够进行接下来的版面分析。而且连续两次查找连通域和绘制连通域的边界盒，

可以合并大量相邻的区域，减少后续算法的运算量。以汉字为例，汉字是象形文字，仔细观察印刷体汉字会发现，大多数汉字是一个连通域，然而很多汉字由几部分组成，这样一个汉字可能就由几个连通域组成。比如“达”字，它由三个连通域组成，但是两次绘制完成就变成了一个连通域；而一些左右或上下明显分开的字形如“北”、“三”等，则有可能被绘制成两个或多个方块。虽然仍有不能合并的文字结构以单独连通域的形式存在，但它们在数量上已经很小，不会对整个算法的运算时间造成明显影响。

两次查找连通域和绘制边界盒的过程，蕴含了连通域合并的原则在其中：如果任意两个连通域边界盒存在重叠部分则他们被一个恰好能包含它们的新连通域边界盒所代替，可以用如下公式表示。

如果对于连通域 1、2 的边界盒，满足

$$\text{Max}(\text{right1}, \text{right2}) - \text{Min}(\text{left1}, \text{left2}) < \text{width1} + \text{width2} \quad (3-1)$$

$$\text{Max}(\text{bottom1}, \text{bottom2}) - \text{Min}(\text{top1}, \text{top2}) < \text{height1} + \text{height2} \quad (3-2)$$

则用新边界盒来替换原来的两个边界盒。

$$\text{left} = \text{Min}(\text{left1}, \text{left2}) \quad (3-3)$$

$$\text{right} = \text{Max}(\text{right1}, \text{right2}) \quad (3-4)$$

$$\text{top} = \text{Min}(\text{top1}, \text{top2}) \quad (3-5)$$

$$\text{bottom} = \text{Max}(\text{bottom1}, \text{bottom2}) \quad (3-6)$$

其中，left、right、top、bottom 为连通域边界盒的位置坐标，width，height 为连通域边界盒的宽度与长度。合并的原理示意图如图 3-5 所示。

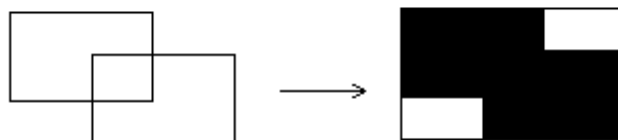


图 3-5 连通域合并示意图

Fig.3-5 The Combination of connected Domains

对示例图像进行抽取的初始连通域结果，如图 3-6 所示

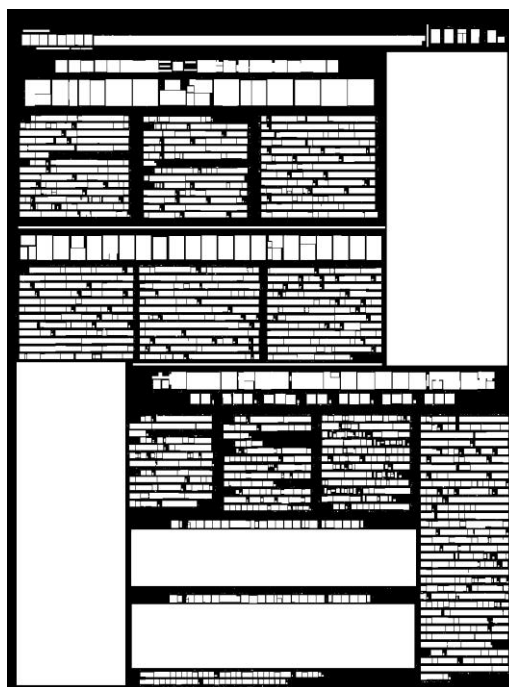


图 3-6 初始连通域

Fig.3-6 Original connected domains

同样由于连续两次查找连通域和绘制连通域的边界盒，对相邻的区域合并，在后续的操作中，将只有一次连通域绘制的机会，否则就会使不应该合并的文字区域合并，导致不能够正确地对嵌入式矩形版面进行正确的分析。

边界盒绘制,将使文档图像的基本要素矩形化。由于文字和直线的结构比较紧密、规则,不会产生附带的不良后果;对于图像和表格来说,如果图表形状不规则,并且有嵌入图表中的文字,这些文字也将被划分到图表的区域中。因此本文算法不适用于图文混排等非矩形版面结构的文档图像。而读书机器人的阅读对象几乎不会涉及到这种类型的版面,因此这个瑕疵是可以接受的。

3.4.2 直线、图表和文字的区分

连通域的组成要素有这样几类:直线、表格、图像、带框文本、文字。版面分析完成后,直线、表格、图像应当被区分开来,予以标记;带框文本和文字,应当被划分成逻辑独立的一个个文本块,并分别予以标记。现在,由于带框文本的属性与图像、表格一致,而文本块则需要在区分出文字之后进一步处理得到,若要一步完成版面分析,将面临复杂的规则和大量的运算。所以在这一节,我们只把连通域划分为直线、图表和文字,并且记录它们的位置、尺寸、属性信息,对示例图像分析的结果如图 3-7 所示;进一步的版面分析将在 3.5 节完成。

1. 直线:直线的长宽比例悬殊(远大于 1 或者远小于 1),我们使用 w/h 来代表连通域宽度和高度的比值,则可以使用公式表示为:

$$(w/h \gg 1) \cup (h/w \gg 1) \quad (3-7)$$

2. 图表:由于图表通常带有边界,或者具有大块的近似颜色,带有图表元素的连通域通常相对较大。我们使用 W 、 H 代表连通域的宽度和高度,则可以使用公式表示为:

$$(w > W/a) \cap (h > H/b) \cup (w \times h > W \times H/c) \quad (3-8)$$

这里, a, b, c 是经验值

3. 文字:仅含文字要素的连通域,不具备以上特征。

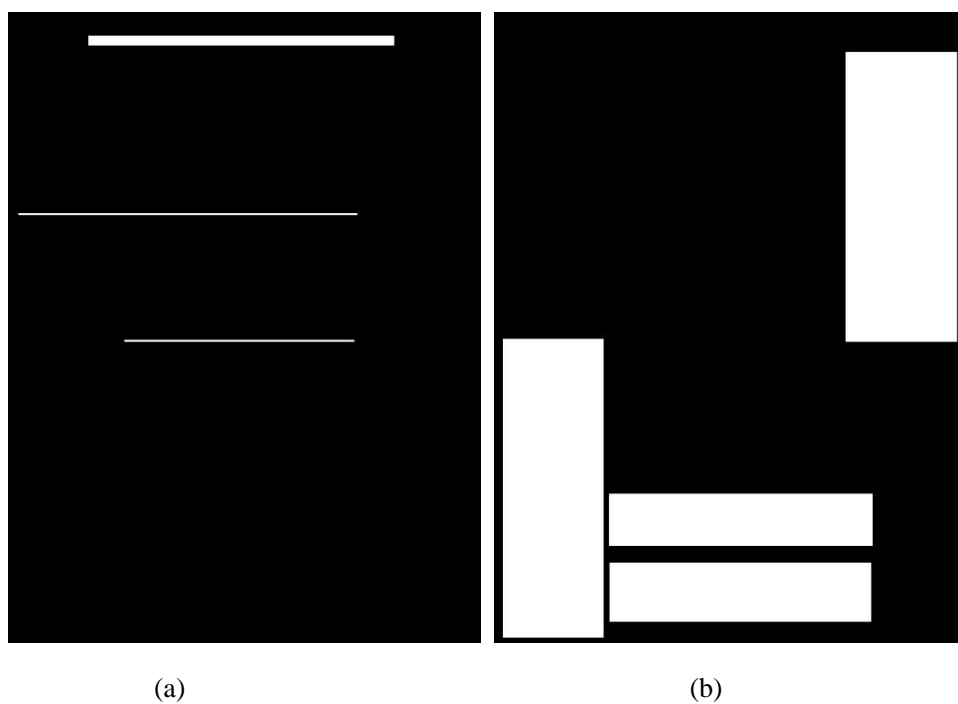


图 3-7 基本结构分类结果

(a) 直线连通域; (b) 图表连通域; (c) 文字连通域

Fig.3-7 Basic Layout Element Classification Result

(a) Line domains; (b) Chart domains; (c) Text domains

3.5 版面结构细分

经过版面基本结构分类，我们获得了各连通域位置信息（连通域矩形框左上顶点坐标），尺寸信息（连通域矩形框的长宽），以及初步属性信息（组成要素是直线、图表还是文字？）。在这一节将对这些信息进一步细化，为阅读顺序确定做好准备。

对于直线，直接留作阅读顺序确定的信息。对于图表，根据它们的纵横投影，可以区分为图像、表格和带框文本；带框文本在去框后，作为单独图像进行分析，获取其中文字区域；对于文字，通过数学形态学的方法，获取逻辑独立的最大文本块，并记忆它们的位置、尺寸、属性信息。

3.5.1 图表区域的分析

根据图表区域的位置、尺寸信息，找到它们在二值缩小图像中对应的位置，然后对相应的区域进行纵横投影。接下来根据纵横投影结果，就可以区分图像、表格和带框文本。

1. 表格

对于表格区域的判定比较有效的方法是通过纵向和横向投影搜索特征线，即表格的框线。现在不少的表格并不具备完整的纵横框线，因此，只要横向或者纵向投影之一具有规律的尖峰，并且尖峰的跨度较小，就可以判定为表格。搜索表格的特征线是认定表格区域的重要实现手段。

一个具有完整框线的表格纵横投影如图 3-8 所示。事实上，可以根据表格线的信息提取表格结构，进一步提取每个单元格的文字信息；只不过对于本文所服务的读书机器人 JoyTon 没有阅读表格的需求，所以在这里就不进行详细的阐述了。

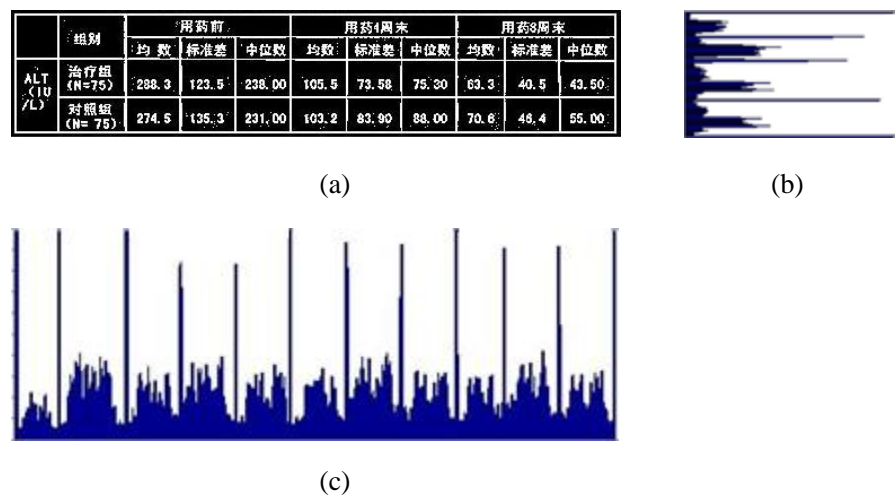


图 3-8 表格区域的纵横投影

(a) 表格区域; (b) 表格区域的水平投影; (c) 表格区域的垂直投影

Fig.3-8 Horizontal and Vertical Projection Histograms of Tables

- (a) A Table;
- (b) Horizontal Projection Histogram of the Table;
- (c) Vertical Projection Histogram of the Table

2. 带框文本

对于带框文本区域的判定比较有效的方法是对外框的特征进行认定。带框文本的外框将在纵横投影的两侧表现为小跨度的尖峰，而且纵横投影之一会存在有规律的较大跨度波谷。通过搜索纵向和横向投影，如果得到外框特征，就可以判断为带框文本。

一个带框文本区域的纵横投影如图 3-9 所示。根据外框特征，可以确定外框对应原图的位置，保存外框的位置、尺寸信息，留作阅读顺序确定使用。同时，提取带框文本内部的子文档图片，把它作为单独的文档图片处理，直到划分为直线、表格、图片、外框、文字为止。直线、表格、图片、外框的位置、尺寸信息保存，作为阅读顺序确定的依据；文字则需要进一步的提取逻辑独立的最大文本块。

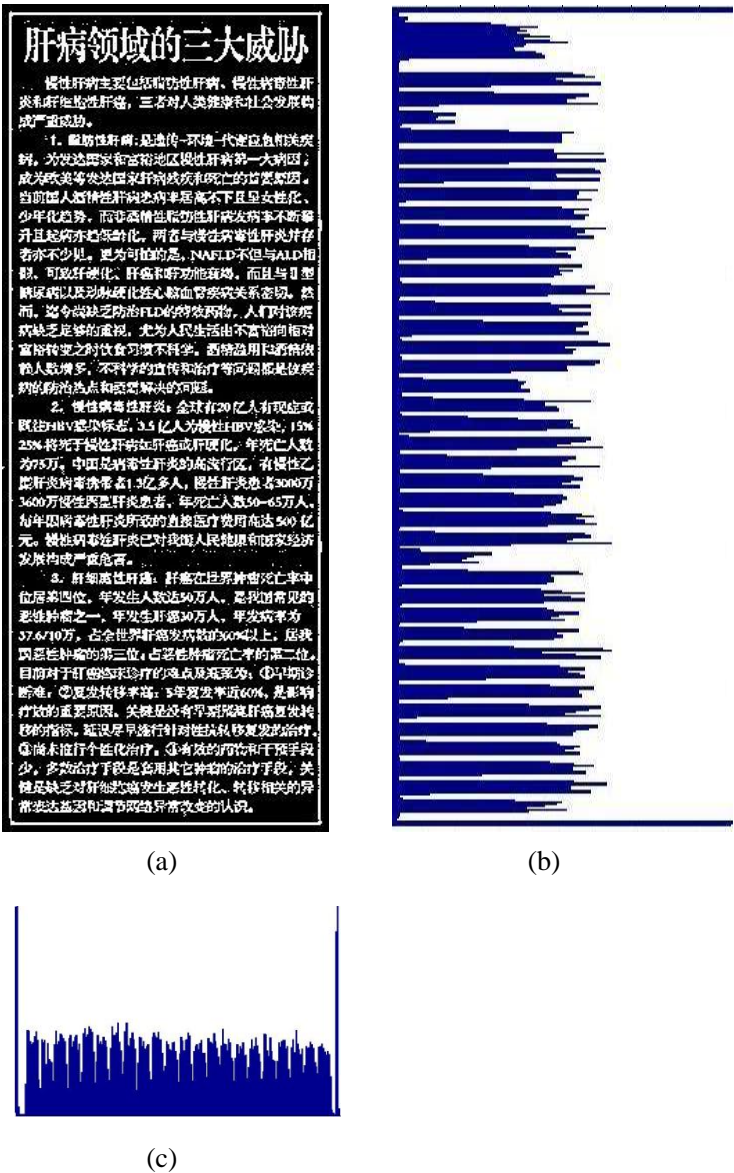


图 3-9 带框文本区域的纵横投影

(a) 带框文本区域； (b) 带框文本区域的水平投影； (c) 带框文本区域的垂直投影

Fig.3-9 Horizontal and Vertical Projection Histograms of Frame-texts

- (a) A Frame-text;
- (b) Horizontal Projection Histogram of the Frame-text;
- (c) Vertical Projection Histogram of the Frame-text

3. 图像

图像的线条通常不具备水平或者垂直的特征，因此图像连通域的纵横投影通常不具备明显的特征。带框图片也与带框文本不同：带框文本的外框将在纵横投影的两侧表现为小跨度的尖峰，而且纵横投影之一会存在有规律的较大跨度波谷；带框图片则仅仅具有外框在纵横投影的两侧表现出的小跨度尖峰。实验证明，把没有特征作为图片的特征，可以相当准确地判断图片区域的属性。

3.5.2 文字区域的分析

为了减少阅读顺序确定过程中的运算量，避免过度频繁调用光学字符识别对极小的文本块进行处理，我们需要进行文字区域的分析，得到逻辑独立的最大文本块，即文本段落或者相同属性的文本段落集，而不是单个汉字这样的小块。这样，只需要对少量相对关系比较简单的逻辑独立文本块进行阅读顺序判定即可，大大减少了判定规则的复杂程度，而且由于计算对象减少，运算时间也会相应大幅减少，这对于读书机器人来说，可以减少等待时间，增加阅读的流畅度。

在预处理和版面分析前期处理过程中，本文通过对图像进行缩小和微膨胀，以及连通域的两次搜索和外切矩形的两次绘制，将字的内部及字之间的间隙进行了有效的填充，因此在这里，文字区域的主要表现形态是两个标点、空格或者换行符间的短句连通域。

如果采用单纯的自底向上的版面分析思路，需要进行逐级合并，最终得到大的连通域。这样的方法处理速度仍然太慢，因此本文采用数学形态学的方法来继续填充各连通域之间的空隙。膨胀和腐蚀算子，将根据文字连通域的统计信息自动获取，以提高算法对不同字体、字号、版面结构的适应性。

3.5.2.1 基本数学形态学运算

数学形态学是以形态结构元素为基础对图像进行分析的数学工具。它的基本思想是用具有一定形态的结构元素去度量和提取图像中的对应形状以达到对图像分析和

识别的目的。数学形态学的数学基础和所用语言是集合论，因此它具有完备的数学基础。

数学形态学是由一组形态学的代数算子组成的，它的基本运算有 4 个：膨胀、腐蚀、开操作和闭操作，它们在二值图像和灰度图像中各有特点。基于这些基本运算还可推导和组合成各种数学形态学实用算法，用它们可以进行图像形状和结构的分析及处理，包括图像分割、特征抽取、边界检测、图像滤波、图像增强和恢复等。本文中通过对二值图像进行数学形态学运算膨胀、腐蚀，来合并文字连通域，使初始分散的文字连通域成为逻辑独立的最大文本块。

运算 1：膨胀

由于 A 和 B 是 Z^2 中的集合，A 被 B 膨胀定义为：

$$A \oplus B = \{z | (\hat{B})_z \cap A \neq \emptyset\} \quad (3-9)$$

这个公式是以得到 B 的相对于它自身远点的映像并且由 z 对映像进行位移为基础的。A 被 B 膨胀，是所有位移 z 的集合，这样， \hat{B} 和 A 至少有一个元素是重叠的。根据这个解释，也可以定义为：

$$A \oplus B = \{z | [(\hat{B})_z \cap A] \subseteq A\} \quad (3-10)$$

集合 B 通常被称为膨胀算子。

如图 3-10 所示，X 是被膨胀的对象，B 是膨胀算子，对于任意一个在阴影部分的点 a，Ba 击中 X，所以 X 被 B 膨胀的结果就是阴影部分。对实际的文字图像进行膨胀的效果如图 3-11 所示。

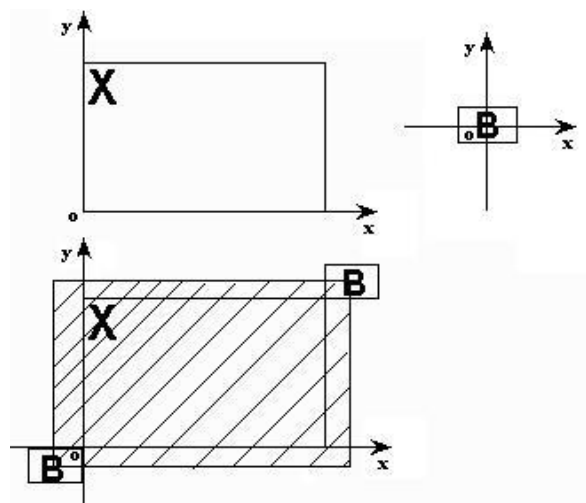


图 3-10 X 区域膨胀的示意图

Fig.3-10 Dilation of X

**Hi,I'm phoenix .
Glad to meet u.**

(a)

**Hi,I'm phoenix .
Glad to meet u.**

(b)

图 3-11 膨胀实例

(a) 原图; (b) 膨胀结果

Fig.3-11 Dilation Sample

(a) A Sample Image; (b) The Dilation Result;

运算 2：腐蚀

可以看做是膨胀(dilation)的对偶运算，对 Z^2 中的集合 A 和 B ， A 被 B 腐蚀定义为：

$$A \ominus B = \{z | (B)_z \subseteq A\} \quad (3-11)$$

这个公式表明，使用腐蚀算子 B 对集合 A 进行腐蚀就是所有 B 中包含于 A 中的点 z 的集合用 z 平移。

图 3-12 中显示了与图 3-10 相似的过程。 X 是被处理的对象， B 是结构元素。对于任意一个在阴影部分的点 a ， B_a 包含于 X ，所以 X 被 B 腐蚀的结果就是那个阴影部分。阴影部分在 X 的范围之内，且比 X 小。对同一幅文字图像进行腐蚀的效果如图 3-13 所示。

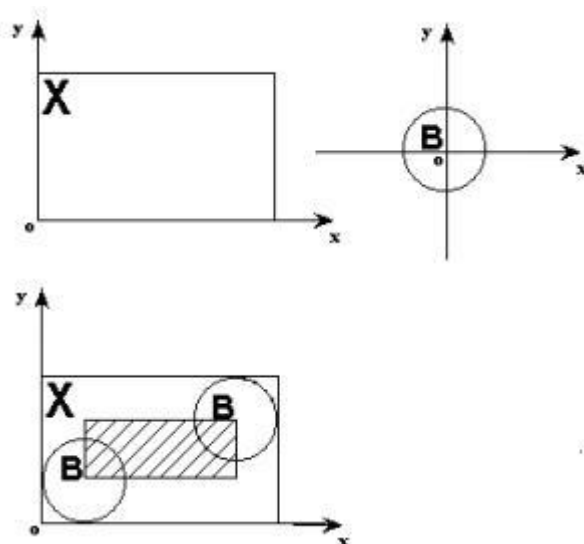


图 3-12 X 区域腐蚀的示意图

Fig.3-12 Erosion of X

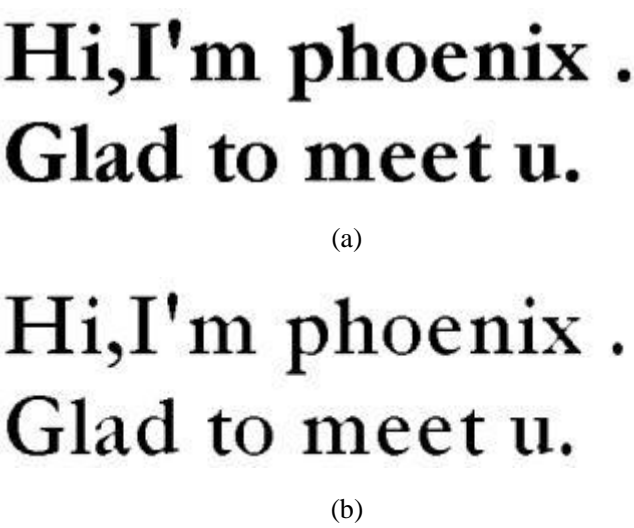


图 3-13 腐蚀实例

(a) 原图; (b) 腐蚀结果

Fig.3-13 Erosion Sample

(a) A Sample Image; (b) The Erosion Result

运算 3：开操作

使用同一个算子先腐蚀后膨胀称为开操作(open)，定义为：

$$A \circ B = (A \ominus B) \oplus B \tag{3-12}$$

开操作一般使对象的轮廓变得光滑，断开狭窄的间断和消除细的突出物。对同一幅文字图像进行开操作的效果如图 3-14 所示。



图 3-14 同一幅图像开操作结果

Fig.3-14 Open Operation Result

运算 4：闭操作

使用同一个算子先膨胀后腐蚀称为闭操作(close),

$$A \cdot B = (A \oplus B) \ominus B \tag{3-13}$$

闭操作同样使轮廓线更光滑，但与开操作相反，它通常消除狭窄的间断和细长的鸿沟，消除小的空洞，并填补轮廓线中的断裂。对同一幅文字图像进行闭操作的效果如图 3-15 所示。



图 3-15 同一幅图像闭操作结果
Fig.3-15 Close Operation Result

3. 5. 2. 2 使用闭操作获取逻辑独立的最大文本块

由于不同的文档中文字的字体、字号不同，初始文字连通域之间的距离不同，因此通过闭操作获取逻辑独立最大文本块的关键是，如何根据当前文档图片中文本的字体字号自动获取合适的膨胀腐蚀算子。通常情况下，文档图片中包含的正文文本数量远大于标题（大字体），标注（小字体）的数量，因此本文采用的方法是，统计所有文字连通域的高度，把去掉高度最大、最小的 10% 后的平均高度，作为文档图片中正文文本的行高 H ，公式如下。并依据这个有当前文档图片统计信息获得的行高，计算膨胀腐蚀算子。

$$H = \frac{\sum h_i}{0.8 \times N} \tag{3-14}$$

其中， H 表示正文文本行高的统计值； h_i 表示高度最大、最小的 10% 除外的文字连通域高度； N 为文字连通域的个数。

现在，我们考察文档的版面性质。中文文档一般是从左到右的横向排版，纵向排版多见于标题，而且水平相邻汉字的间距比较小。字体字号相同的情况下，一个标点符号占用宽度相当于 $1/4$ 到 $3/4$ 个行高 H 。格式相同的上下相邻两行的行间距约为 $3/4$ 个行高（单倍行距情况下）。在多列文本格式中，相邻两个文本块的间距约为 $3/2$ 个行高。而英文文档一般是从左到右的横向排版，段落中相邻文字、文字行的间距通常小于中文文档。因此定义膨胀腐蚀算子 B 为：

$$B_w = \frac{H \times 2}{2} + 1 \quad (3-15)$$

$$B_h = \frac{\frac{3}{4} \times H \times 2}{2} + 1 \quad (3-16)$$

其中， B_w 为膨胀算子 B 的列数， B_h 为膨胀算子 B 的行数， H 为正文文本行高的统计值。

使用这个膨胀腐蚀算子对示例图像的文字连通域进行闭操作，得到逻辑独立的最大文本块如图 3-16 所示。

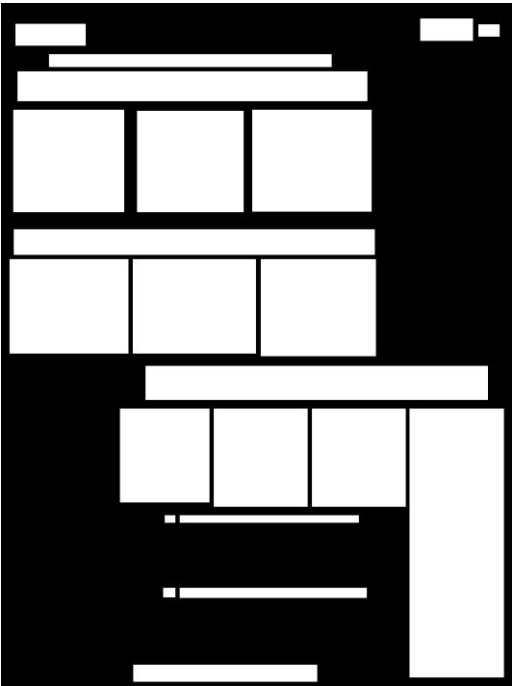


图 3-16 逻辑独立的最大文本块

Fig.3-16 Maximum homogeneous text domains

通过文字区域的数学形态学运算，初始分散的文字连通域得到合并，成为逻辑独立的最大文本块。这些逻辑独立的最大文本块，分别由相同属性的文字段落构成，在逻辑上可以作为个体看待；同时，最大程度上合并了相同属性的文字段落，降低了阅读顺序判定的对象数量和规则复杂程度。

3.6 版面分析的成果

本章通过对要素特征的理解和版面几何结构的分析，从文档图像中提取连通域的位置、尺寸信息，并把它们区分为文本、图像、表格、线框，然后对文本连通域进行数学形态学操作，划分为逻辑独立的最大文本块。各版面要素的属性、位置、尺寸信息，将用于阅读顺序确定；逻辑独立的最大文本块，将是阅读顺序确定的操作对象，

它们的位置、尺寸信息不仅用于接下来的阅读顺序确定，而且将用来切割子图片进行光学字符识别。

第四章 文本块的阅读顺序确定

4.1 阅读顺序确定的目的

版面通常具有几何结构和逻辑结构。几何结构包括区域的划分及其位置、大小等；逻辑结构则蕴含了各组成部分之间的相互关系。版面分析的任务是识别出图像中具有相似属性的区域，解析其几何结构，而版面理解则是在版面分析的基础上，确定版面的逻辑关系，即以人对该版面的理解为基础，依赖于特定的类型、特定领域的知识，判断版面各区域对象间的逻辑关系^[31]。

一般版面可分为基于文本的版面和基于图形的版面两类。在前一类版面中，文本占主要部分，如报纸、杂志等，结构化较强；在后一类版面中，图像或图形占主要部分，如地图、画报、设计图等，文本仅起辅助作用。提取基于文档的版面逻辑结构称为文档版面理解；提取基于图形的版面逻辑结构称为图形版面理解。本文主要讨论文档版面理解中文本对象的顺序确定方法，通过解读版面规则中蕴含的逻辑关系，对逻辑独立的纯文本区域进行排序。

由于日常书报杂志具有各种版面结构，从而导致各个文本块之间逻辑关系复杂，如果不进行逻辑解读，而直接由读书机器人 JoyTon 阅读，将会出现几篇文章掺杂阅读，不能正常表达文章内容的情况。对于盲人和视力障碍者而言，这种情况将会妨碍他们理解日常书报杂志，不能获取需要的信息，更无法享受阅读的乐趣。

在设定版面的时候，出版社都会考虑到方便读者阅读的因素，因此，日常书报杂志的几何结构和逻辑结构之间的对应关系还是会遵循一些普适的原则，这些原则是简单而且固定的，可以根据它们进行比较准确的版面逻辑结构解读。例如它们相对严格地遵守着西方和中文传统的自上而下、从左向右的排版习惯；例如根据字体的相对大

小，可以知道哪些区域代表文章的标题和正文。这一点，是本文阅读顺序确定算法的现实基础。

本章将在对阅读顺序确定算法的研究现状进行回顾之后，提出一种分层的递归阅读顺序确定算法。这种算法是基于纵横切割法的自顶向下算法，最大的特点是运算速度快；另外，由于对线框信息的利用和补充规则的添加，能够基本保证对日常书报杂志排序的准确性。

4.2 阅读顺序确定的研究现状

版面分析和理解的技术已有几十年的历史，国内外许多研究机构，都在版面分析与理解领域建立了研究中心，并且取得了一定的研究成果。目前，版面分析和理解技术的研究，已经可以对不同领域的版面，采用不同的技术进行研究，并逐步得到实际应用。

版面分析是版面几何结构的生成过程，而版面理解则是将版面的几何结构映射为逻辑结构的过程。同样的版面布局并非表示同样的逻辑含义，而具有同一逻辑含义的版面可以有多种布局方式。因此，从几何结构到逻辑结构的映射，不是简单的一一对应关系。目前多数版面理解方法是，将几何对象直接映射为逻辑对象，这只适用于已知的、简单的版面布局与逻辑结构，对于较为复杂的版面尚无能为力。另外，在研究中，对版面几何结构生成方法的讨论较多，而对于版面理解的注意则较少。

通常情况下，由于版面的种类繁多复杂，不同种类的版面具有不同的排版规范，同类的版面也会有细节上的变化，使得版面几何结构和逻辑结构之间的对应关系非常复杂，难以建立统一的转换模型，进行准确的解读。为了降低两者对应关系的复杂度，通常采用两种方法。一种是把目标版面类型限制在某一类结构化版面内，例如，名片，信函，科技论文^[32]等。这些版面的几何结构和逻辑结构对应关系比较明确，可以直接通过建立版面模型，对各版面要素间的逻辑关系进行解读。另一种方法是，将与版面相关的知识引入系统^[33, 34]，指导版面要素间逻辑关系的分析。而阅读顺序确定的算法，大致可以分为如下三种：

基于树形分割结果的算法^[35]，利用递归纵横分割方法，直接导出各个不同区域间的顺序关系。将整个文档版面看作根节点，将第一次分割的结果作为它的子节点；再依次对各个子节点进行分割，直到不能分割为止。这种方法运算简单，实现迅速。

基于格式描述语言的算法^[36]，将版面的几何结构和逻辑结构都用一系列矩形区域来描述，将知识规则都用形式定义语言（Form Definition Language, FDL）来表示，使用程序自动解析 FDL 的涵义，分析版面中各区域间的相互关系。基于格式描述语言的算法，是最有效的解读版面逻辑结构的方法之一；但是如何完整且有效地使用 FDL 定义逻辑结构隐含的知识，如何设计程序，是目前存在的主要问题。

基于知识的算法^[37]，使用不同的版面模型来定义不同版面的先验知识，在使用这些模型将版面表示成可分析的结构，从而解读版面的逻辑结构。其中，最重要的是确定各种版面的几何结构和逻辑结构之间的关系，即先验知识。例如，对于书籍、杂志等规则的版面，其排版通常是有一定规律可循的，利用这些规律，可以对版面分析的结果进行逻辑结构解读。

4.3 阅读顺序确定的算法流程

为了避免对每种文档版面结构先验信息的参考，本文参考 J.L. Meunier 的方法^[33]，提出了一组简单的规则，通过这些描述了日常书报杂志在版面结构上普适原则的规则，本文基于纵横切割法，建立了一个递归的方法来作为阅读顺序确定算法的主体。由于直线、方框具有区域划分的意义，所以在进行主体操作之前，先利用第三章获得的线框信息，将文本区域进行划分、排序。由于文本中会存在一些特殊的结构，不能够使用纵横切割法进行分割，因此对于它们，单独建立规则来进行阅读顺序确定。具体的操作流程如图 4-1 所示。

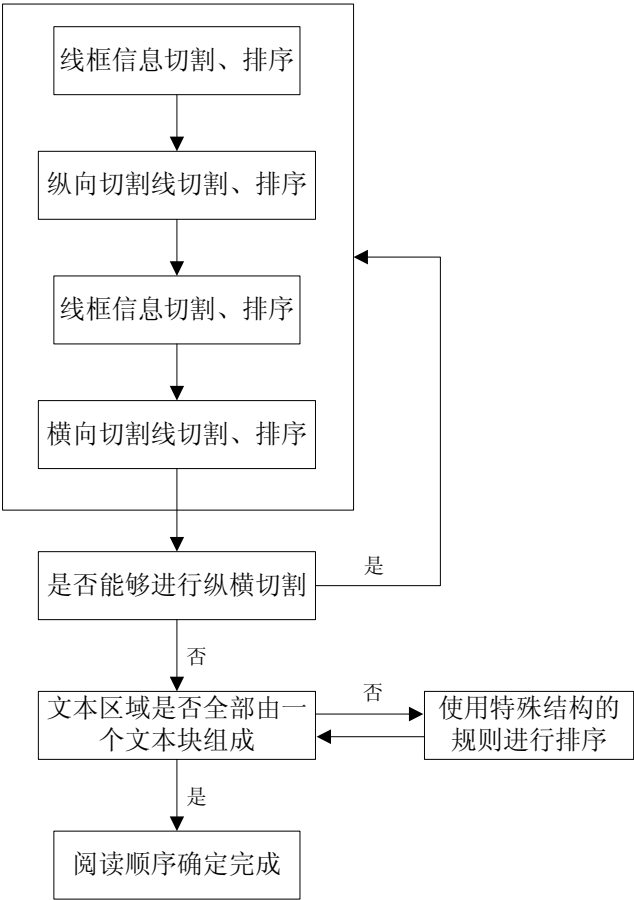


图 4-1 阅读顺序确定流程图

Fig.4-1 The Workflow of Reading Order Determination

4.4 阅读顺序确定算法

4.4.1 排版规则

日常书报杂志的版面排版规则相对而言比较简单而且统一，归纳起来主要有这样几种：

1.分栏规则

一篇文章通常被划分为多个栏，属于同一篇文章的各栏通常具有相同的宽度，而相邻文章栏的宽度则可能不一样。对于横排版文本，同一栏的各文本块在相同的列上排列。同一文章栏各文本块的宽度和左右坐标应是相似的，只有当图像或标题区域嵌入文本区域时，受影响的文本块才会出现变形。

2. 排版次序规则

通常遵守自上而下、从左向右的排版次序规则。

3. 标题规则

标题一般字号大、行数(或列数)少，出现在正文的上方、左侧、右侧或嵌入中间；位置在水平、垂直两个方向的某个中轴附近；主标题字体比副标题大，副标题靠近主标题，标题与文本间距偏大；竖排版标题一般在正文的右侧。

4. 段落规则

文本段落通常作两端对齐的排列，如果有标题或插图嵌入文本区域时，文本在剩余的区域内作两端对齐的排列；段落的常规格式是：首行为缩进、悬垂或不满一行。

5. 图表说明规则

图表都有相应的图说明和表头。图说明的位置一般在图的下方，与正文分开，此时图与版面内容相关，若无图说明，则为装饰版面用图，与版面内容无关，表头一般在表格的上方，分一行或多行，与正文分开。

6. 版面附加信息规则

主要指页眉和页脚，页眉通常处于版面的正上方，与正文以一条横贯版面宽度的直线分开，上有一行文本或小图标；页脚在版面下方，通常只有页码，其位置有三种：居中、左对齐和右对齐。

根据标题、正文之间的排版规律，及其他文本段的位置关系，可以方便地判断出文本段的标题、页眉、页脚等，但是目前读书机器人 JoyTon 中没有进行这样的实现，因此，本文不把重点放在标题、页眉、页脚的区分，而是集中在各个文本块顺序确定这一点。

4.4.2 基于树结构的阅读顺序确定

为了确定版面对象的顺序，这里针对一般的日常书报版面，建立空间结构树，通过遍历树得到版面对象顺序。这一节将按照对文档图像处理的顺序，首先将整个文档图像作为根节点，利用第三章获得的直线、方框信息将文本区域划分、排序；然后将每个文本区块作为根节点，基于描述了日常书报杂志在版面结构上普适原则的规则，采用纵横切割的方法，进行递归排序，直至不能分割为止；最后检验上一步获得的文本块，如果全部由一个逻辑独立的最大文本块组成，则阅读顺序确定完成，否则，说明文档中存在特殊结构，不能够使用纵横切割法进行分割，因此对于它们，采用单独建立的规则来进行阅读顺序确定。

4.4.2.1 利用线框信息的排序

由于直线、方框具有区域划分的意义，这是出版社编辑们在设定排版的时候，为帮助读者版面理解而给出的信息，因此，它们在区域分割上的优先级别是最高的，我们首先利用线框信息，将文本区域划分、排序。这一步操作，在提高阅读顺序确定准确性的同时，提高了算法的效率。

对于带框文本，方框明确地把它定位成一个单元，因此，在接下来的递归纵横分割操作中，相对整个文档图像，带框文本是一个不可分割的单元；而带框文本内部，同时进行纵横分割，解析其内部结构。

直线是明确的区域分割信号，直线贯通文档图像的比例越大，它的分割强度越大。对于横向和纵向的直线，分别按照如下公式计算每条直线贯通文档图像的比例，按照比例由大到小的顺序，直线的分割强度依次降低。对于横向直线，以其位置的纵坐标为分界点，划分为上区域、下区域，依据日常书报杂志自上而下的通用排版规则，代表上区域的子节点优先于代表下区域的子节点；对于纵向直线，则利用其横坐标为分界点，分为左区域和右区域，代表左区域的子节点先于代表右区域的子节点。

$$R = \frac{L_w}{W} \quad (4-1)$$

$$R = \frac{L_h}{H} \quad (4-2)$$

其中， R 表示直线贯通文档图像的比例， W 、 H 表示文档图像的宽高， L_w 、 L_h 分别表示横向、纵向直线的长度。

4.4.2.2 利用普适原则排序

根据排版规则，我们可以注意到这样两点：一是在中国大陆使用的日常书报，不论是中文的，还是英文的，还是中英文混排的，都会遵从西方传统的自左向右由上而下的横向排版顺序，即使在报纸中出现少量的竖排版，也是存在于标题中，可以看作是一列只有一个文字的横向排版；二是不论一个文档被划分为何种规格的栏，栏间距总是大于行间距，而纵向的分栏尤其常用。主要依据这两个方面，本文基于纵横切割法，建立了一个自上而下的递归方法作为阅读顺序确定算法的主体。

Nagy、Seth 和 Viswanathan 在 1992 年提出了一个经典的版面分析的算法——纵横切割法^[14]，它以文字图像在二维空间水平方向和垂直方向的投影函数为基础，通过分析函数曲线的走向，实现文档结构的分析，是典型的空间域的自顶向下算法。

为了方便理解，首先定义几个版面对象间的关系：

1. 包含关系：

若有版面对象 S 和 A ，且在空间上对象 S 包含对象 A ，则版面对象 S 、 A 之间为包含关系，在版面逻辑关系树上，对象 S 是对象 A 的根节点，对象 A 是对象 S 的一个子节点。

2. 上下关系：

若有版面对象 A 和 B ，在空间上对象 A 在对象 B 的上侧；对象 A 、 B 具有共同的根节点 S ，且两个对象之间除了分隔符无其他版面对象，则版面对象 A 、 B 之间为

上下关系。根据日常书报杂志自上而下的排版规则，对象 A、B 为兄弟节点，且对象 A 优先于对象 B。

3. 左右关系：

若有版面对象 A 和 B，在空间上对象 A 在对象 B 的左侧；对象 A、B 具有共同的根节点 S，且两个对象之间除了分隔符无其他版面对象，则版面对象 A、B 之间为左右关系。根据日常书报杂志从左向右的排版规则，对象 A、B 为兄弟节点，且对象 A 优先于对象 B。

可以使用纵横切割法进行分割的文本区域之间，文本区域与逻辑独立最大文本块之间，以及逻辑独立最大文本块之间的关系，都可以使用如上定义的三种关系来描述。

首先，对于文本区域，计算区域图像的垂直投影，并检测所有可能的纵向切割线。所谓纵向切割线，是指能够把文本区域切割成两个部分，但是不会穿过任何逻辑独立最大文本块的纵向贯穿线，表现在区域图像的垂直投影上，就是一个波谷。假如检测到 n 条纵向切割线，就能够把文本区域切分成 $n+1$ 个子区域，子区域之间符合左右关系，初始区域与子区域之间符合包含关系。按照左右关系的定义，可以对这 $n+1$ 个子区域进行排序。

如果子区域由不止一个逻辑独立的最大文本块组成，那么这个子区域还有继续切分的可能。对于有继续切分可能的子文本区域，计算水平投影，并检测所有可能的横向切割线。所谓横向切割线，是指能够把文本区域切割成两个部分，但是不会穿过任何逻辑独立最大文本块的横向贯穿线，表现在区域图像的水平投影上，就是一个波谷。假如检测到 n 条横向切割线，就能够把文本区域切分成 $n+1$ 个子区域，子区域之间符合上下关系，初始区域与子区域之间符合包含关系。按照上下关系的定义，可以对这 $n+1$ 个子区域进行排序。

因为日常书报杂志是横向排版，通常为纵向分栏，因此纵向切割线的优先级大于横向切割线。如果横向切割线的数量大于 1，那么可能只有部分横向切割线是必要的，例如图 4-2 所示的情况，下侧的一条横向切割线是不必要的，而且会造成阅读顺序确定出错。因此我们必须判断哪些横向切割线是必要的，并将不必要的横向切割线造成

的影响去除。检验所有新的子区域是否共享同一条纵向切割线，如果有相邻的子区域共享纵向切割线，那么我们合并这两个或者多个区域。

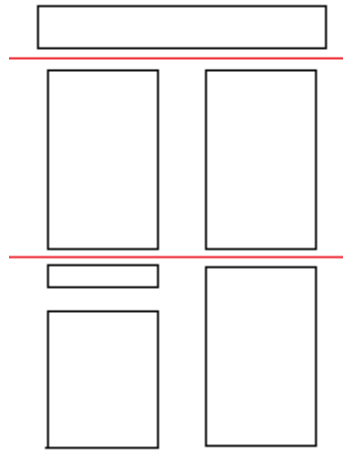


图 4-2 多余横向切割线示意图

Fig.4-2 Excess Horizontal Cutting Line

反复进行纵向切割线和横向切割线的搜索，直到所有文本区域都不能够进行纵横切割为止。这个时候，可能还存在一个区域当中包含多个逻辑独立文本块的情况，但是基于纵横切割法的递归阅读顺序确定无法对这样的结构进行处理，因此，在下一节中，我们将详细讲述，如何对包含特殊结构的文本区域进行阅读顺序确定。

4. 4. 2. 3 特殊结构内的阅读顺序确定

一些特殊的版面结构，例如图 4-3 所示的 L 形结构、嵌入式矩形结构等，不能够使用纵横切割法进行正确的区域分割。这种情况在书籍和技术文档中并不多见，因此很多版面理解的算法没有涉及这个领域，然而在报纸杂志当中版面结构更加复杂，出现特殊版面结构的频率也随之提高，这是读书机器人 JoyTon 需要涉及的领域，因此本文需要针对这些特殊结构给出专门的规则，使阅读顺序确定过程能够顺利地进行。

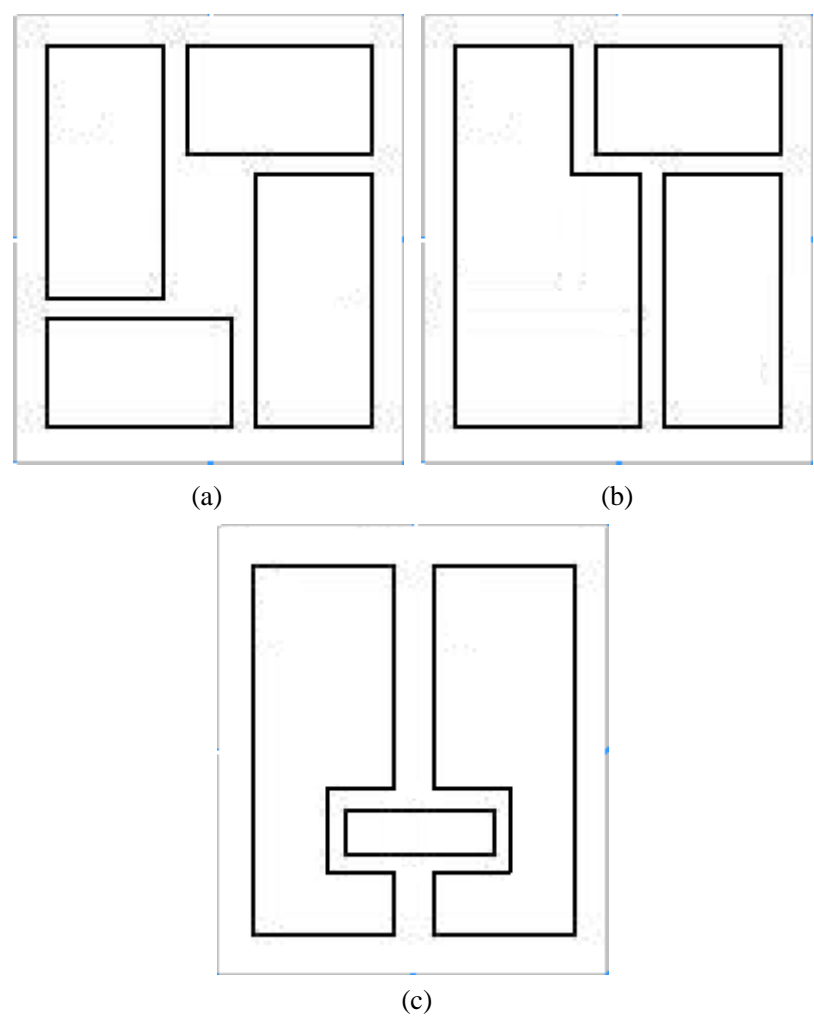


图 4-3 特殊版面结构
(a) L 形 1; (b) L 形 2; (c) 另一种特殊版面结构

Fig.4-3 Special Layout Configurations

(a) L-shape 1; (b) L-shape 2; (c) Another Special Layout Configuration

在对特殊结构进行研究的过程中发现，特殊结构所包括的逻辑独立文本块之间的联系，相对于整个文档的其他部分更加紧密，也就是说，与文档的其余部分相比，特殊结构可以看作一个整体。因此，在进行纵横切割、排序的过程中，把特殊结构作为一个子节点进行处理和排序，符合版面逻辑结构的要求。

在对特殊结构进行研究的過程中还发现，虽然特殊结构内部的逻辑结构与几何结构对应关系非常复杂，但是各个逻辑独立文本块所表达的信息却是相对独立的。对于这些文章段落，阅读顺序的变化，往往不会影响对于语义的理解；而语义逻辑严密的特殊结构，又会更加严格的遵守自左向右由上而下的排版规则。因此，本文对特殊结构的规则进行简化，使用各个逻辑独立文本块的左上顶点坐标（位置信息）来代替文本块整体，作为阅读顺序判断的依据，具体方法如下：

- 1. 把特殊结构整体作为根节点；
- 2. 计算逻辑独立文本块左上顶点坐标与该特殊结构左上顶点坐标之间的距离；
- 3. 将距离最近的逻辑独立文本块作为第一子节点，其余部分作为第二子节点；
- 4. 对第二子节点采用纵横切割排序的方法进行阅读顺序确定；
- 5. 检验所有叶子节点是否由一个单独的逻辑独立文本块组成，如果是，阅读顺序确定过程完成；否则，把包含多个逻辑独立文本块的子节点作为新的特殊结构进行处理，直到所有叶子节点都由一个单独的逻辑独立文本块组成。

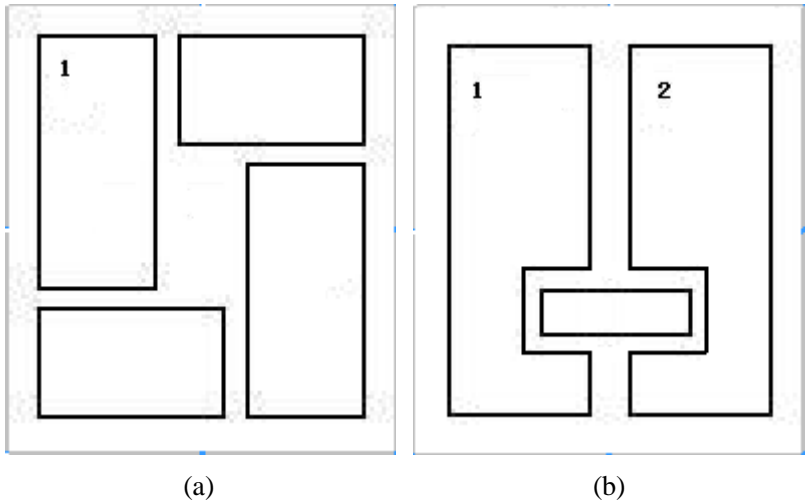


图 4-4 提取文本块

(a) L 形； (b) 另一种特殊版面结构

Fig.4-4 Extraction of Text Domain

(a) L-shape; (b) Another Special Layout Configuration

提取文本块的结果如图 4-4 所示。对于图 4-4a，提取标记为 1 的文本块后，就能够完成阅读数字确定；对于图 4-4b 则需要进行两次循环，提取标记为 1、2 的两个文本块之后，才能完成特殊结构内部的阅读顺序确定。

4.5 文本块的阅读顺序确定

4.4 节分三个层次描述了本文的阅读顺序确定算法，对各个局部进行了比较详尽的描述；本节则总结阅读顺序确定算法，按照如下所述的八个步骤进行操作，就可以将文档图像的几何结构转换为逻辑结构，用于读书机器人的阅读。

1. 建立一个根节点，代表整幅文档图像；将带框文本看作一个逻辑独立文本块。如果一个节点所代表的文本区域只有一个逻辑独立的文本块，这个节点就是最终叶子节点；否则，这个节点为兴趣节点。
2. 如果代表整幅文档图像的根节点是最终叶子节点，则跳转到第 7 步；否则检验所有没有使用过的直线区域。如果一条直线贯穿一个兴趣区域，则这条直线被标记为使用过，并且在被贯穿区域的节点下，创建两个子节点，左侧（或者上侧）区域的子节点优先于右侧（或者下侧）区域的子节点。
3. 检验所有节点，如果没有兴趣节点，则跳转到第 7 步；否则，计算所有兴趣节点所代表区域的垂直投影，检测所有可能的纵向切割线。如果有 n 条纵向切割线，则在该兴趣节点下，创建 $n+1$ 个子节点，按照左右关系的规则，确定所有子节点的优先顺序。
4. 检验所有节点，如果没有兴趣节点，则跳转到第 7 步；否则，计算所有兴趣节点所代表区域的水平投影，检测所有可能的横向切割线。如果有 n 条横向切割线，则在该兴趣节点下，创建 $n+1$ 个子节点，按照上下关系的规则，确定所有子节点的优先顺序。

如果横向切割线的数量大于 1，则检验所有新的子区域是否共享同一条纵向切割线。如果有相邻的子区域共享纵向切割线，那么我们尝试合并这两个或者

- 多个区域；如果合并后的区域有纵向切割线，那么我们创建一个新的节点来代表合并后区域，并且使用它来替代被合并区域的所有节点。
5. 检验所有节点，如果没有兴趣节点，则跳转到第 7 步；否则，对于所有兴趣节点，跳转到第 2 步，直到不能进行纵横切割为止。
 6. 检验所有节点，如果没有兴趣节点，则跳转到第 7 步；否则，这个兴趣节点代表的区域具有特殊结构。比较同一个兴趣区域内所有逻辑独立文本块的左上点的位置，找到最接近兴趣区域左上点的逻辑独立文本块，并创建两个子节点，优先的一个来代表最左上位置的文本块，另一个代表剩余的文本区域；对于剩余的文本区域，回到第 2 步，直到不再存在兴趣区域为止。
 7. 把代表带框文本的节点作为根节点，重复第 2 到第 6 步。
 8. 现在所有的叶子节点都是最终叶子节点，因此遍历整棵有序树就可以确定最终的阅读顺序。



图 4-5 阅读顺序确定的结果

Fig.4-5 Result of Reading Order Determination

对版面分析所使用的示例文档图像进行阅读顺序确定的结果，如图 4-5 所示，读书机器人 JoyTon 按照如图所示的序列阅读所有的文本块，就可以正确地表达报纸所包含的信息了。

5、实验结果与分析

本文的目标是，通过对文档图像的版面分析和阅读顺序确定算法进行研究，提取和利用文档的版面信息，解析文档的要素和规则，提升读书机器人阅读日常书报杂志的能力。因此，版面分析和阅读顺序确定算法具有较高的正确率，速度以及对不同语言文字不同版面的适应能力，是保证读书机器人流畅工作的必要条件，也是对算法优劣考察的重点。

本章在第一节首先给出对一个报纸文档进行版面分析和阅读顺序确定的处理过程和结果，展示本文算法的准确度。第二节则是给出了 200 张文档图像进行版面分析和阅读顺序确定的统计结果，并把正确率和运算时间与其他算法进行比较，得出本文算法具有更高正确率、运行速度以及适应能力的结论。

5.1 对某一张报纸文档的处理结果

图 5-1 展示了对一张报纸文档进行版面分析和阅读顺序确定的过程。

图 5-1(a)是一个初步版面分析得到的单页文档图像。图 5-1(b)是图像增强、二值化、图像扭曲校正后的结果，这幅图像去除了噪声和扭曲，文字清晰，适合进行后续图像处理和光学字符识别。图 5-1(c)是进行图像缩小和微膨胀后的结果，可以看到，相邻的文字已经连接在一起，大大减少了搜索到的连通域数量，可以减少后续运算的复杂度。图 5-1(d)是连通域搜索的结果，可以看到直线、图表、文字连通域表现出明显不同的几何特征；图 5-1(e, f, g)分别是提取出的直线、图表、文字连通域。图 5-1(h)是对文字连通域进行数学形态学操作，得到的逻辑独立的最大文本块。图 5-1(i)是对各个文本块进行阅读顺序确定的结果，读书机器人只需要按照编号顺序依次阅读，就可以帮助盲人或视力障碍者获取报纸信息，帮助他们享受阅读的乐趣了。

中华肝病医学杂志 2014年第12期

中国人民解放军第三〇二医院杨永平教授研究指出

控制肝损伤是治疗肝病的关键

杨永平教授指出，肝损伤是肝病发生发展的关键环节。控制肝损伤是治疗肝病的关键。肝损伤的病理机制复杂，涉及多种因素，包括病毒、免疫、代谢、环境等。在临床实践中，应针对不同病因采取相应的治疗策略，以减轻肝损伤，延缓疾病进展。

六味五灵片给慢性乙型肝炎肝损伤治疗带来新曙光

六味五灵片作为一种新型中药制剂，在慢性乙型肝炎肝损伤的治疗中展现出显著优势。该药具有清热解毒、活血化瘀、保肝护肝等功效，能够有效减轻肝损伤，改善肝功能，提高患者生活质量。临床研究表明，六味五灵片联合抗病毒治疗，能显著提高治疗有效率，减少副作用。

肝病治疗的五大误区

杨永平教授指出，肝病治疗中存在五大误区，患者应避免：1. 盲目追求“特效药”，忽视规范治疗；2. 轻信偏方，延误病情；3. 过度治疗，增加经济负担；4. 忽视生活方式调整，影响治疗效果；5. 缺乏长期随访，导致病情反复。

中华肝病医学杂志 2014年第12期

中国人民解放军第三〇二医院杨永平教授研究指出

控制肝损伤是治疗肝病的关键

杨永平教授指出，肝损伤是肝病发生发展的关键环节。控制肝损伤是治疗肝病的关键。肝损伤的病理机制复杂，涉及多种因素，包括病毒、免疫、代谢、环境等。在临床实践中，应针对不同病因采取相应的治疗策略，以减轻肝损伤，延缓疾病进展。

六味五灵片给慢性乙型肝炎肝损伤治疗带来新曙光

六味五灵片作为一种新型中药制剂，在慢性乙型肝炎肝损伤的治疗中展现出显著优势。该药具有清热解毒、活血化瘀、保肝护肝等功效，能够有效减轻肝损伤，改善肝功能，提高患者生活质量。临床研究表明，六味五灵片联合抗病毒治疗，能显著提高治疗有效率，减少副作用。

肝病治疗的五大误区

杨永平教授指出，肝病治疗中存在五大误区，患者应避免：1. 盲目追求“特效药”，忽视规范治疗；2. 轻信偏方，延误病情；3. 过度治疗，增加经济负担；4. 忽视生活方式调整，影响治疗效果；5. 缺乏长期随访，导致病情反复。

(a)

(b)

中华肝病医学杂志 2014年第12期

中国人民解放军第三〇二医院杨永平教授研究指出

控制肝损伤是治疗肝病的关键

杨永平教授指出，肝损伤是肝病发生发展的关键环节。控制肝损伤是治疗肝病的关键。肝损伤的病理机制复杂，涉及多种因素，包括病毒、免疫、代谢、环境等。在临床实践中，应针对不同病因采取相应的治疗策略，以减轻肝损伤，延缓疾病进展。

六味五灵片给慢性乙型肝炎肝损伤治疗带来新曙光

六味五灵片作为一种新型中药制剂，在慢性乙型肝炎肝损伤的治疗中展现出显著优势。该药具有清热解毒、活血化瘀、保肝护肝等功效，能够有效减轻肝损伤，改善肝功能，提高患者生活质量。临床研究表明，六味五灵片联合抗病毒治疗，能显著提高治疗有效率，减少副作用。

肝病治疗的五大误区

杨永平教授指出，肝病治疗中存在五大误区，患者应避免：1. 盲目追求“特效药”，忽视规范治疗；2. 轻信偏方，延误病情；3. 过度治疗，增加经济负担；4. 忽视生活方式调整，影响治疗效果；5. 缺乏长期随访，导致病情反复。

中华肝病医学杂志 2014年第12期

中国人民解放军第三〇二医院杨永平教授研究指出

控制肝损伤是治疗肝病的关键

杨永平教授指出，肝损伤是肝病发生发展的关键环节。控制肝损伤是治疗肝病的关键。肝损伤的病理机制复杂，涉及多种因素，包括病毒、免疫、代谢、环境等。在临床实践中，应针对不同病因采取相应的治疗策略，以减轻肝损伤，延缓疾病进展。

六味五灵片给慢性乙型肝炎肝损伤治疗带来新曙光

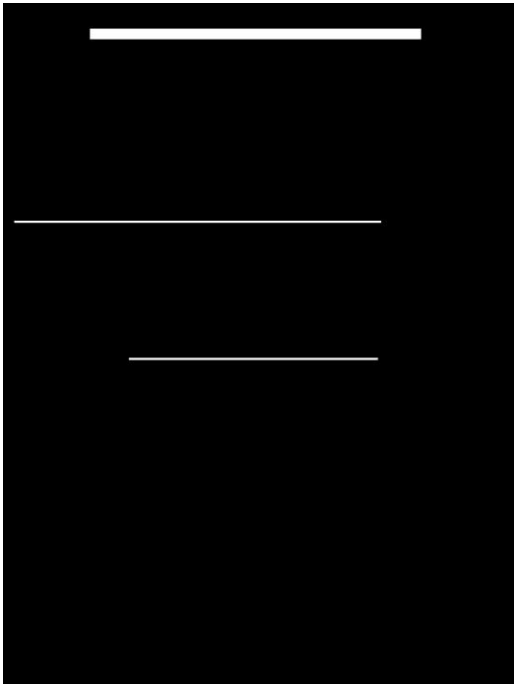
六味五灵片作为一种新型中药制剂，在慢性乙型肝炎肝损伤的治疗中展现出显著优势。该药具有清热解毒、活血化瘀、保肝护肝等功效，能够有效减轻肝损伤，改善肝功能，提高患者生活质量。临床研究表明，六味五灵片联合抗病毒治疗，能显著提高治疗有效率，减少副作用。

肝病治疗的五大误区

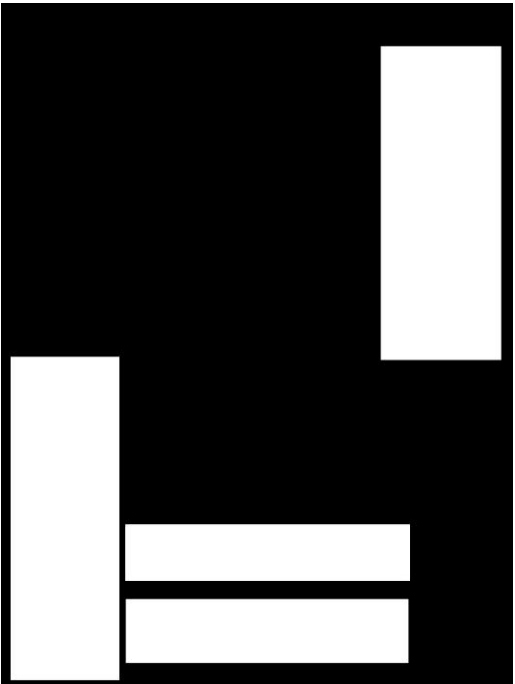
杨永平教授指出，肝病治疗中存在五大误区，患者应避免：1. 盲目追求“特效药”，忽视规范治疗；2. 轻信偏方，延误病情；3. 过度治疗，增加经济负担；4. 忽视生活方式调整，影响治疗效果；5. 缺乏长期随访，导致病情反复。

(c)

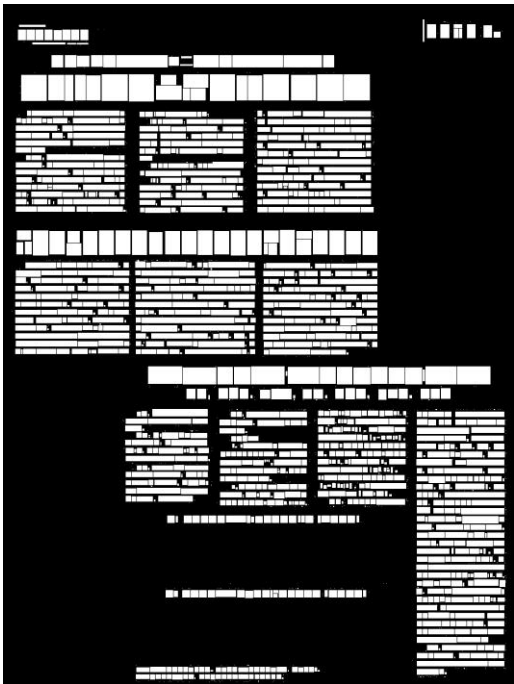
(d)



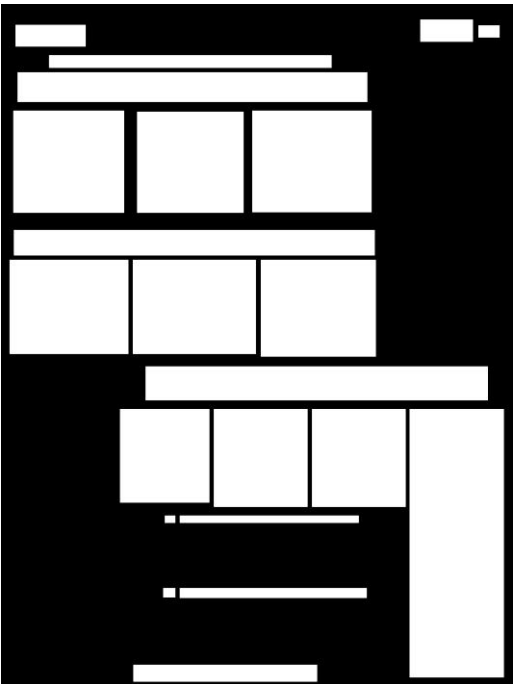
(e)



(f)



(g)



(h)



图 5-1 一张报纸文档图像的处理过程

Fig.5-1 The Process of a Newspaper Document Image

5.2 与其他算法的比较

本文的版面分析和阅读顺序确定算法，已经被应用到大量文档图像上。这里，我们使用一个包含了 200 幅典型文档图像的样本来进行实验，它们中有中文文档，也有英文文档；有 50 张书籍图片，50 张科技杂志图片，100 张普通杂志或者报纸图片。实验环境是佳能相机 A620，抓取图像大小为 1600×1200，pc 的处理器为 Intel 奔腾 IV-2GHz，内存为 1.25G，系统开发环境 Visual Studio 2005。

本文的方法和钟辉等的版面分析和重构方法^[38]，目标都是日常书报杂志，而本文的方法，采用了分层的结构，提出了更好的区分表格、图片的方法，而且采用了完全不同的获取逻辑独立最大文本块的方法。表 5-1 是两种方法的比较，可以看到，本文的方法不论是准确程度还是运算时间，都要比钟辉等提出的方法更有优势。对于这 200 幅典型文档图像组成的样本，具有 96.4%的准确率，同时也说明了，本文的版面分析算法具有对不同版面结构的良好适应能力。

表 5-1 200 幅文档图像样本的版面分析结果及算法比较

		原始图像	本文算法	钟辉等的算法
区域数量	表格区域	67	66	66
	图片区域	180	177	152
	文字区域	956	917	590
区域正确分类率			96.4%	67.2%
平均运行时间			0.131s	0.143s

本文的阅读顺序判定方法参考了 J.L. Meunier 的方法，因此表 5-2 给出两种算法的比较。在本文的方法中，不仅正确率得到了一定的提升，达到 96%，而且在运算流畅上，也不再出现由于存在特殊结构而死循环，无法完成阅读顺序确定的情况。J.L.

Meunier 的阅读顺序确定算法，之所以无法计算时间，是因为在遇到 L 形结构时会陷入死循环，或者不能够得到最终的阅读顺序确定结果。如果排除这个因素，J.L. Meunier 的方法平均耗时 0.029s，只略小于本文的方法；而牺牲 0.002s 的时间，换取准确度和适应性的提高是值得的。

表 5-1 200 幅文档图像样本的阅读顺序确定结果及算法比较

	本文算法	Meunier 的算法
原始图像数量	200	200
正确排序的图像数量	192	183
正确排序的图像比例	96%	91.5%
平均运行时间	0.031s	不可测量

从上面的实验结果中，我们看到在面对变化多样的版面结构的情况下，本文提出的方法仍然有效地提高了正确率。版面分析过程中，造成区域划分错误的主要原因是过度切割。例如，图 5-2 是一个由少量文字组成的文本区域，可能会由于标点、空格造成的空隙而被划分为两个文本区域。这样的错误，会增加调用光学字符识别的频率，增加一定的计算量，也有可能影响到语义逻辑的表达。

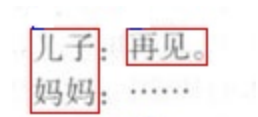


图 5-2 版面分析错误的主要原因
Fig.5-2 Main Reason of Layout Analysis Errors

6、总结与展望

6.1 本文主要研究工作及成果

本文的主要研究内容在于读书机器人 JoyTon 的机器视觉模块，包括机器人获取图像后的预处理、版面分析和理解三个部分，着重对文档图像的版面分析和阅读顺序确定算法进行研究，通过提取和利用文档的版面信息，解析文档的要素和规则，提升读书机器人阅读日常书报杂志的能力。具体来说，包括以下几个部分：

1) 初步版面分析

本文中针对特定的背景条件和环境，先对读书机器人获取的文档图像进行初步版面分析，提取单页的无背景文字图像，算法效果明显。

2) 图像增强和二值化算法

本文采用了经典的图像锐化算法对读书机器人中获取的文字图像进行增强。由于读书机器人获取的文字图像所处的光线环境不确定,常常会有背景色渐变的情况出现，并且会面临光线整体比较暗的情况，本文参考 Sauvola 的算法进行二值化。

3) 文字图像的几何扭曲校正算法

通过参考以往学者的工作，并针对特定的工作环境，提出了一种适用于读书机器人的文字图像的几何扭曲校正算法。首先分析文字图像的连通域，定位文字行，然后用三次样条曲线来拟合文字行，最后根据顶部和底部的两条曲线作为基准行来复原成无扭曲的图像。因为在本算法中建立数学模型的依据是文字图像上检测出来的文字行，所以数学模型适用性好，鲁棒性较强，实际的试验效果也比较理想。

4) 图像缩小和微膨胀

由于版面分析并不需要文字、图表等要素的细节纹理，而图像缩小可以大量减少版面分析的运算量，所以在预处理阶段进行一步图像缩小的操作，并且采用基于块结构的缩小算法，使得图像缩小过程中，图像的前景进行少量的膨胀，这个膨胀，可以将紧邻的文字连通域合并。

5) 文档图像的版面分析

版面分析是纸介质信息数字化的首要步骤，也是本文两个重点之一，其目的在于提高数字化处理的智能程度，通过对要素特征的理解和版面几何结构的分析，从文档图像中提取逻辑独立区域的位置信息，并把它们区分为文本、图像、表格、线框，为阅读顺序确定提供各区域位置、属性等相关信息。

本文中首先扫描预处理获得的二值缩小图像，抽取文档图像上所有的连通域，然后根据连通域的尺寸特征，区分直线、图表和文字。对于直线，留作阅读顺序确定的信息。对于图表，根据它们的纵横投影，可以区分为图像、表格和带框文本；带框文本在去框后，作为单独图像进行分析，获取其中文字区域；对于文字，通过数学形态学的方法，获取逻辑独立的最大文本块，并记忆它们的位置属性信息，为阅读顺序确定做准备。

6) 文本块的阅读顺序确定

本文在前人版面理解研究的基础上，专注于文本区域阅读顺序确定的工作，提出一种分层的递归阅读顺序确定算法。这种算法基于描述了日常书报杂志在版面结构上普适原则的规则，采用纵横切割的方法，进行递归排序，直至不能分割为止；在利用普适原则之前，利用直线、方框信息将文本区域分成区块辅助排序；另外在遇到特殊结构时，如 L 形结构，不能使用纵横切割法进行排序，则建立相应的规则来进行阅读顺序确定。

最后，通过实验证实，本文的版面分析方法正确率为 96.4%，阅读顺序确定方法正确率为 96%，平均运算时间为 0.162s，从而保证了读书机器人稳定，流畅的工作，帮助盲人和视力障碍者阅读日常书报杂志。

6.2 结论与展望

本文中实现的图像预处理、版面分析和阅读顺序确定算法，基本上可以满足读书机器人 JoyTon 的正常工作，但是还是有许多可以改进的部分，读书机器人的视觉模块，还有很多后续工作等待展开。

1. 图像几何扭曲校正和版面分析的平衡

本文研究中发现,对于图像扭曲比较严重的多栏复杂版面,同时需要进行图像几何扭曲校正和版面分析,但是两者之间相互依赖,只有另一方取得较好的成果时,才能进行有效的图像处理。对于这个问题,本文的处理是在要求读书机器人正确摆放文档进行拍照,并且获得图像扭曲较小的情况下进行的。由于在盲人和视力障碍者使用的过程中,不能够像普通人一样保证文档摆放的位置良好,在后续的研究中,需要对图像几何扭曲校正和版面分析的平衡进行专门的考量,以减少对文档摆放位置的要求。

2. 标题和说明文字的处理

根据标题、正文之间的排版规律,图表和相应说明文字间的位置关系,及其他文本段的位置关系,可以方便地判断出文本段的标题、页眉、页脚、图表说明文字。虽然本文没有把重点放在这个部分,但是这些信息的单独说明会帮助听众理解文档中包含的信息,因此也具有研究价值。

3. 具有特点的文档版面库的补充

本文的研究建立在普通日常书报杂志的普遍规律基础上,但是日常生活中总会遇到具有特殊版面结构的文档,比如名片、账单等,因此,应当总结中文特殊版面的特点与规律,增加名片、账单、表格、公式的版面结构辅助信息,提高相应算法的效率与准确度,使读书机器人能够自动识别相应文档,并快速准确的进行分析和阅读。

4. 彩色图书的处理

本文假定获取的文字图像是灰度图,没有考虑读书机器人朗读彩色图书时的情况,换言之,就是忽略了图书的颜色信息。事实上,图书的颜色信息通常包含了丰富的数据^[39],所以彩色文字图像预处理也将是本课题后续研究的方向之一。

5. 人机交互

本文中并没有涉及使用者给予读书机器人反馈的情况,但是,往往会出现正在阅读的段落并不是使用者感兴趣的部分,或者发生了阅读顺序判断错误的情况,如果可以

由使用者选择当前逻辑独立的最大文本块跳过或者延后的话,可以更好的起到帮助盲人和视力障碍者享受阅读乐趣的作用。

目前,作者所在的实验室正在更深入地研究读书机器人,不论对硬件还是软件都在进行进一步的优化。希望在以后的研究中,能够不断优化流程,提升性能,可以更好地服务盲人,造福盲人,并且为相关技术的发展抛砖引玉,为服务弱势群体的橙色关怀提供更好的技术支持。

参 考 文 献

- [1] 赵群飞, 吴心然, 唐矫燕. 自动翻书机. 中华人民共和国.发明专利, 授权专利: ZL 200410017692.7, 2004.4.
- [2] 冶建科, 赵群飞, 周毅, 张伟业. 一种自动翻页扫描机器人的研制. 技术应用. 2009, 01: 27-31.
- [3] 唐矫燕, 赵群飞, 杨汝清, 吴心然. 读书机器人机构设计. 上海交通大学学报. 2005, 39(12): 2025-2028.
- [4] Dumitras T., Lee M., Quinones P., Smailagic A., Siewiorek D., Narasimhan P. Eye of the Beholder: Phone-Based Text-Recognition for the Visually-Impaired. Wearable Computers, 2006 10th IEEE International Symposium on. 2006: 145-146.
- [5] V. Gaudissart, S. Ferreira, C. Thillou, B. Gosselin. Mobile Reading Assistant for Blind People. Proc. 13th European Signal Processing Conference, EUSIPCO. 2005.
- [6] Mori S., Suen C.Y., Yamamoto K. Historical Review of OCR Research and Development. Proc. IEEE. 1992, 80(7): 1029-1057.
- [7] Govindan V.K., Shivaprasad A.P. Character Recognition a Review. Pattern Recognition. 1990, 23(7): 671-683.
- [8] 张炳中. 汉字识别技术. 北京:清华大学出版社. 1992.
- [9] Khan A., Khan S., Mahmood W. MRST: A New Technique for Information Summarization. Transactions on Engineering, Computing and Technology V4 February. 2005: 1305-5313.
- [10] George Nagy. Twenty Years of Document Image Analysis in PAMI. IEEE Trans. on Pattern Analysis and Machine Intelligence. 2000, 22(1): 38-82.
- [11] T. Akiyama, N. Hagita. Automated Entry System for Printed Documents. Pattern Recognition. 1990, 23(11): 1141-1154.
- [12] 陈自利. 基于小波与神经网络的文字识别系统研究[博士论文]. 重庆: 重庆大学. 1999.
- [13] 姜哲, 夏莹. 中文版面分析技术. 第六届全国汉字识别学术会议论文集. 1996: 131-136.
- [14] G. Nagy, S. Seth, M. Viswanathan. A Prototype Document Image Analysis System for Technical Journals. Computer. 1992, 7(25): 10 - 22.
- [15] R. Gonzalez, R.Woods. Digital Image Processing, 2. 北京:电子工业出版社.2007:59-113.
- [16] 贺志明. 数码相机拍摄的文本图像的二值化. 上海工程技术大学学报. 2006, 20(3): 214-218.

- [17] W. Niblack. An Introduction to Digital Image Processing. Prentice-Hall, Englewood Cliffs, New Jersey.1986: 115-116.
- [18] J. Sauvola, T. Seppanen, S. Haapakoski, M. Pietikainen. Adaptive Document Binarization. Proc. of the 4th Int. Conf. on Document Analysis and Recognition. 1997: 147-152.
- [19] 张伟业. 读书机器人中的图像预处理算法设计与开发[硕士论文].上海:上海交通大学. 2010.
- [20] 靳从. 中文版面分析关键技术的研究[博士论文]. 南京: 南京理工大学. 2007.
- [21] 贾娟, 陈堃铎, 周东浩. 图文互斥版面中文字阅读顺序的确定. 中文信息学报. 2005, 19(5): 67-75.
- [22] 贾娟, 陈堃铎. 多篇章非 Manhattan 版面层次布局模型和阅读顺序确定. 计算机辅助设计与图形学学报. 2005, 17(2): 372-377.
- [23] S. Lee, D. Ryu. Parameter-Free Geometric Document Layout Analysis. Pattern Analysis and Machine Intelligence. IEEE Transactions on. 2001, 23: 1240-1256.
- [24] Jain A K, Zhong Y. Page Segmentation Using Texture Analysis. Pattern Recognition. 1996, 29(5): 743-770.
- [25] F. Shafait, D. Keysers, T. M. Breuel. Performance Evaluation and Benchmarking of Six Page Segmentation Algorithms. Pattern Analysis and Machine Intelligence, IEEE Transactions on. 2008, 30:941-954.
- [26] 徐兆军, 业宁, 王厚立. 基于神经网络的版面分析. 计算机应用. 2004, 24:274-275.
- [27] D.Sylwester, S.Seth. A Trainable, Single-Pass Algorithm for Column Segmentation. Proc.3rd ICDAR, Montreal, Canada. 1995: 615-618.
- [28] L.OGorman. The Document Spectrum for Page Layout Analysis. IEEE Trans. On PAMI. 1993, 15(11): 1162-1173.
- [29] A. Simon, J. Pret, A. Johnson. A Fast Algorithm for Bottom-Up Document Layout Analysis. IEEE Trans. Pattern Analysis and Machine Intelligence. 1997,19: 273-276.
- [30] Jiming Liu, YuanY Tang, ChingY Suen. Chinese Document Layout Analysis Based on Adaptive Split-and- Merge and Qualitative Spatial Reasoning. Patten Recognition. 1997, 30(8): 1265-1278.
- [31] 贾娟, 元文法, 侯晓辉, 陈堃铎. 基于不规则版面布局模型的区域划分和分区排序算法. 计算机工程与应用. 2003, 30: 51-53
- [32] Y. Ishitani. Document Transformation System from Papers to XML Data Based on Pivot XML Document Method. International Conference on Document Analysis and Recognition. 2003, 1:250-255.
- [33] J.L. Meunier. Optimized XY-Cut for Determining a Page Reading Order. International Conference on Document Analysis and Recognition. 2005, 1: 347 - 351.

-
- [34] M. Ceci, M. Berardi, G.A. Porcelli, D. Malerba. A Data Mining Approach to Reading Order Detection. International Conference on Document Analysis and Recognition. 2007, 2:924-92.
- [35] S.Tsujimoto, H.Asada, Understanding Multi-articled Documents. In Proceedings of International Conference on Pattern Recognition, Atlantic City, New York, June 16-21. 1990:551-556.
- [36] A. Hashizume, P. S. Yeh, A. Rosenfeld. A Method of Detecting the Orientation of Aligned Components. Pattern Recognition Letters. 1986, 4(1): 125-132.
- [37] T. Parsons. Introduction to Compiler Construction. W. H. Freeman and Company. 1992.
- [38] 钟辉, 孙士兰, 刘倩. 中文版面分析和重构. 沈阳建筑大学学报: 自然科学版. 2008, 24(2): 333-336.
- [39] 黄海凌, 刘列根, 张宇. 彩色文档图像的版面分析. 计算机工程. 2008, 34(15): 231-233.

致 谢

岁月如梭，如歌。转眼之间，我已在上海交通大学研究生院度过了两年多难忘的时光。站在毕业的门槛上，回首往昔，奋斗和辛劳，甜美与欢笑，都化成了丝丝记忆，萦绕心间。上海交通大学以其优良的学习风气和严谨的科研氛围教我求学，以其博大包容的情怀胸襟和浪漫充实的校园生活育我成人。值此毕业论文完成之际，我谨向所有关心、爱护、帮助我的人们表示最诚挚的感谢与最美好的祝愿。

本研究及学位论文是在赵群飞教授的亲切关怀和悉心指导之下完成的。从课题选择，到方案制定实施，论文撰写修改，赵老师都始终给予细心的指导和不懈的支持。两年多来，赵群飞教授渊博的专业知识，严谨的治学态度，精益求精的工作作风，诲人不倦的高尚师德，朴实无华平易近人的人格魅力对我影响深远。赵老师不仅授我以渔，而且教我做人，短短两年多的相处，将令我终生受益无穷。在此向赵群飞教授献上我的谢意。

同时，感谢上级党组织对我的关心和教育，并于 2010 年 6 月接受我成为一名光荣的中共预备党员。感谢我的父母，我们温暖的家是我永远的依靠，也是我最大的动力。感谢自动化系的各位老师，两年多来言传

身教，使我从一个门外汉成长为对模式识别初有涉猎的研究生。感谢答辩委员会的各位老师，百忙之中抽出时间来审阅论文，为我指点迷津。

我还要感谢一起度过研究生生活的各位同学——我的室友杨娅婷，我的同门刘宁、黄孝寅、蔡芝阳，我的师兄师姐和师弟们，以及 B0803293 班的所有同学，正是由于你们的热情帮助，不懈支持，我才能克服一个个困难和疑惑，顺利地完论文走向毕业，我们在一起的苦乐酸甜，将铭刻在我心间。愿友谊长存！

最后，再次向所有关心、爱护、帮助我的人们献上最诚挚的感谢与最美好的祝愿。

攻读硕士学位期间已发表或录用的论文

[1] Yucun Pan, Qunfei Zhao, Seiichiro Kamata, Document Layout Analysis and Reading Order Determination for a Reading Robot, IEEE TENCON 2010 Conference, Nov.22-24, 2010, FUKUOKA, Japan. (EI 检索)