

北方工业大学

2015年优秀硕士学位论文集

NORTH CHINA UNIVERSITY OF TECHNOLOGY

2015 OUTSTANDING COLLECTION OF MASTER'S DEGREE

(下册)

研 究 生 院

目 录

下 卷

计算机学院

- 1 基于 DCI 的版权保护核心架构研究与实现 王 帅(1)
- 2 时间序列数据挖掘在 FY-3 MERSI 数据上的应用研究 田 圆(8)
- 3 基于立体视觉的三维重建算法研究 孔德星(20)
- 4 基于连通域的扭曲文本图像校正方法研究与应用 郭正东(30)

电子信息工程学院

- 5 高速低功耗无线传感器网络数据传输技术与设计 高洪磊(38)
- 6 面向 HEVC 的屏幕内容编码相关技术的研究 张 扬(47)
- 7 高精度北斗天线设计 石璐瑶(58)
- 8 国际空间站环控生保系统建模与仿真 王永皎(69)
- 9 高密度客流实时检测系统设计 乔 帅(77)
- 10 光子晶体慢光特性的研究 权 炜(88)
- 11 分布式生物特征识别技术研究 于晓婕(103)
- 12 基于光通信的 LDPC 通用编解码器的实现与研究 张 锋(113)
- 13 基于 Zigbee 的病人呼叫系统的设计 董婉青(127)
- 14 10 位 10MHz SAR ADC 设计 史立韬(140)
- 15 可编程放大器的分析与设计 朱江南(147)

建筑与艺术学院

- 16 对智能冰箱及未来交互性设计发展的研究 祁文静(158)
- 17 北京地区高密度居住区户外活动空间多功能利用研究 王 昭(189)
- 18 张家口近代城市规划建设研究 范家昱(211)

经济管理学院

- 19 基于公共健康需求的公共空间适老性研究 曹 阳(220)
- 20 中国铜工业景气指数与预警系统研究 帅加琴(229)
- 21 我国铝业生产能量流研究 刘实实(235)
- 22 我国铝业生产环境负荷研究 侯思远(249)
- 23 企业技术进步的低碳经济贡献度研究 魏 洁(255)
- 24 基于商业生态系统的产业链协同创新机理研究——以新能源汽车为例 李雅荣(264)
- 25 文化创意企业团队异质性、知识共享与团队绩效关系研究 秦 爽(270)
- 26 基于社会网络分析的建筑工业化利益相关者关系研究 李晓桐(277)
- 27 基于生态学的京津冀地区创新适宜度评价研究 傅建敏(284)
- 28 价值工程在食品冷链物流企业低碳运营中的应用研究 郭晓薇(295)

文法学院

- 29 强制医疗责任保险立法问题研究 钟 蕾(302)
- 30 网络失范言论的刑法规制 曾 磊(308)

理学院

- 31 V 系统在纹理分类中应用 王 俊(315)
- 32 基于光场成像的三维图像特征点提取与匹配 王晓飞(326)
- 33 基于 V-系统的三维模型数字水印新算法 徐燕青(337)
- 34 非线性 Burgers 方程的高精度数值解法 范 磊(348)
- 35 非线性奇异摄动对流扩散方程的高精度数值解法 李雪峥(359)
- 36 几类线性模型中的 Bootstrap 方法及其应用 梅 波(372)
- 37 基于价格条件 VaR 的套保套利研究 杨桂君(383)

马克思主义学院

- 38 张伯苓“公能”教育思想研究 王 钰(391)

基于 DCI 的版权保护核心架构研究与实现

研究生 王 帅

导师 吴洁明

(北方工业大学计算机学院 北京 100144)

摘 要 本文提出了一种基于数字版权标识符(Digital Copyright Identifier, DCI)的版权保护方案,包括三大核心内容:基于 DCI 的数字作品版权管理,包括数字作品登记、合同备案、提取数字内容特征信息,建立数字作品版权元数据库和特征库;基于 DCI 的第三方费用结算和作品使用统计,获得 DCI 编码的数字作品在数字内容运营商平台被消费时,将交易信息反馈给版权登记机构,权利人可通过版权登记机构获取数字作品的消费信息;权利人可向版权管理服务机构提出监测申请,也可对侵权网站进行举报,发现疑似侵权时自动截取证据,并生成监测报告。

本文提出了基于 DCI 的数字版权保护核心架构,设计核心架构的解决方案。将 DCI 发码体系设计成一种分布式的、可扩展的体系结构,这种体系结构不但能有效实现 DCI 的发放,还为推动 DCI 编码成为国际标准创造条件。

关 键 词 DCI; 数字资源标识符; 版权保护; 版权服务

作者简介 王帅,女,1989 年出生,硕士研究生,主要研究方向为软件工程。

E-mail: xinshizp@163.com

1 绪论

文化产业作为新兴产业,自 20 世纪 90 年代起,在世界各国开始迅速发展,许多国家将文化产业作为战略产业发展^[1]。随着网络化、信息化、数字化等技术的发展,给文化产业的发展带来巨大的发展机遇,同时也使得版权保护面临前所未有的挑战。主要体现在:互联网的开放性导致侵权的低成本与无节制和依靠传统的事后纠纷处理机制无法从根本上解决侵权问题。

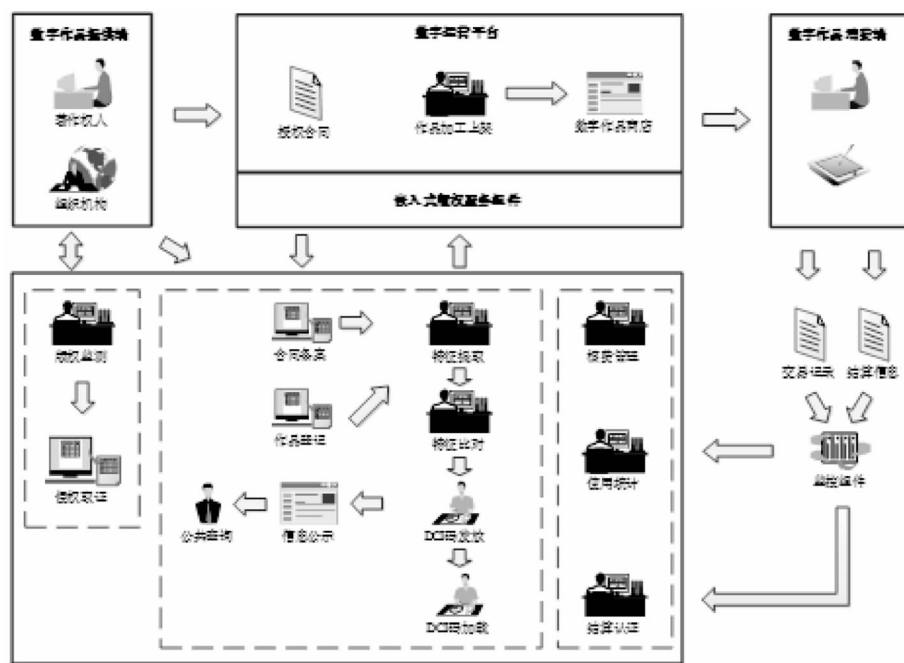
互联网的快速发展给版权产业带来严重的挑战,引发了互联网版权保护的第一次革命,主要特征是将传统的著作权法律制度及管理模式延伸到互联网领域,形成了以“内容保护措施+纠纷处理机制”为基础的版权保护模式。这种模式在初期有效,但无法适应互联网的特性,且存在许多问题^[2]。

随着研究的深入,从根本上解决版权产业链参与者的利益共享,通过技术手段实现快速维权机制,成为版权保护研究的核心命题^[3]。以利益分享为动力和成果,互联网版权保护正在迎来第二次革命^[4]。其主要特征为:突破传统著作权保护制度,通过法律创新,制定更加严格高效的网络版权规制法案;以法律创新为基础,建立起以公共服务为支撑的数字版权保护新模式。

顺应互联网版权保护第二次革命浪潮,建立一种新的版权服务模式,运用版权手段加强内容管理,切实保护数字网络环境下的作品版权,从而有效解决互联网版权保护难题,同时带动版权产业、文化产业发展。此外,建立自主创新的版权服务体系,还有利于增进中外文化产业的交流,保障国家文化领域的安全。

数字版权唯一标识符(Digital Copyright Identifier, DCI)是由中国版权保护中心提出的,对数字作品权利进行唯一标识,可以分配给任何形式的数字作品。DCI 编码由前缀和后缀两部分组成,前缀是 DCI 标识符,后缀是编码详细信息,包括版本号、发码中心编码、分配年、状态标识和分配码几部分。DCI 编码具有唯一性、永久性、兼容性、可维护性和可扩展性等特点。

DCI 是对权利的标识,以 DCI 为基础构建版权保护方案的原理是:著作权人在登记机构进行作品登记和合同备案,登记机构分发 DCI 编码。备案的数字作品在第三方运营平台上架,第三方监控其交易记录和结算信息,进行使用统计和结算认证,实现整个产业链的利益共享。作品提供端可以对登记备案的作品进行侵权监测,由监管机构取证作为诉讼依据,实现快速维权。基于 DCI 的版权保护方案如图 2-1 所示:



数字版权管理服务、数字版权费用结算和侵权监测是数字作品版权保护体系的三大核心功能,建设数字版权管理服务、数字版权费用结算平台和侵权监测平台是版权保护体系的主要任务。其中,数字版权管理服务主要实现作品登记和合同备案,建设版权元数据库和作品特征库;数字版权费用结算平台实现作品登记备案的收费和对第三方运营平台与著作权人费用结算的监控和认证;侵权监测平台为著作权人提供侵权监测和侵权举报两种服务,对网络侵权行为取证。

通过对基于 DCI 的版权保护方案的研究,划分三大平台的业务域和业务组件。在版权保护方案实施时,还应考虑业务系统的划分粒度和划分后的内聚性、耦合性,在业务架构基础上映射版权保护的应用架构,而应用架构的实现是以技术架构为支撑的,解决技术问题才能保证应用架构的实现。版权

保护的 业务架构、应用架构和技术架构分别如图 2—2、2—3、2—4 所示。

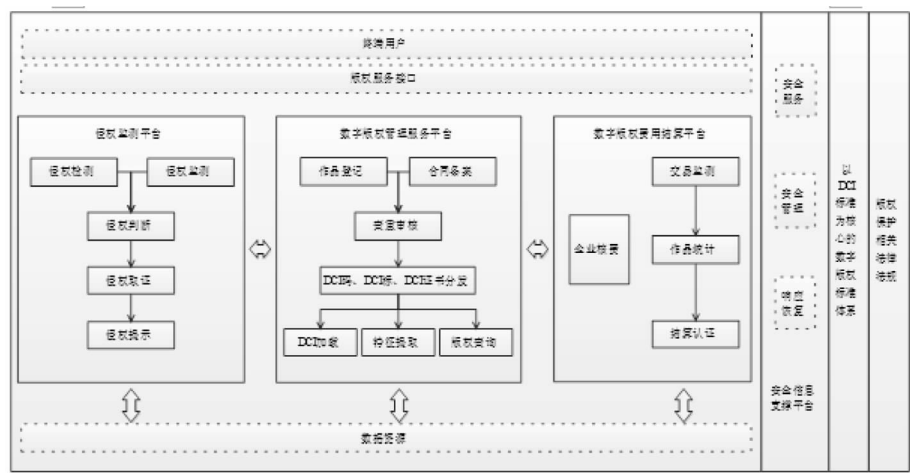


图 2—2 版权保护业务架构

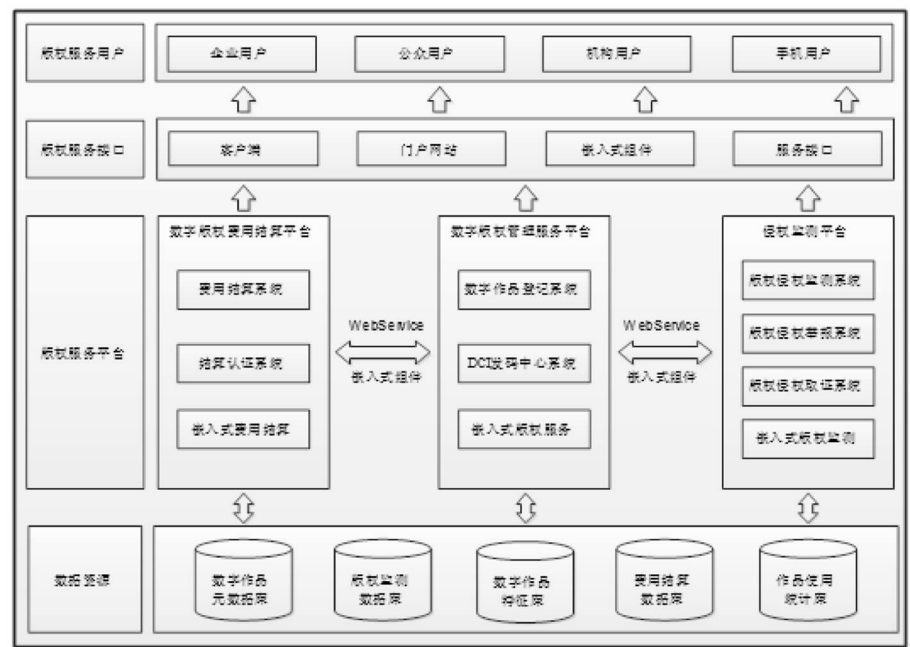


图 2—3 版权保护应用架构

在业务架构、应用架构、技术架构中，全部以 DCI 为主线，以 DCI 分发、解析为基础。DCI 分发和解析是版权保护实现的关键技术，贯穿于整个数字作品版权保护流程，支撑整个数字作品版权保护方案的运行。



图 2-4 版权保护技术架构

2.2 基于 DCI 的版权保护关键技术

DCI 发码、解析是基于 DCI 的数字作品版权保护方案的核心内容,是整个数字作品版权保护方案实现的前提。分析版权产业发展现状,构建通用的体系结构,并考虑随着业务增长,该体系结构应具有一定的可扩展性。针对网络环境下存在的数字作品种类丰富、数量大等问题,应设计快速高效的 DCI 发码和 DCI 解析方案。

1. DCI 发码方案

参照 DCI 的 Handle System 发码体系^[5],DCI 发码体系应构建成为一种分布式、可扩展的结构,应设置 DCI 管理中心、DCI 发码中心、版权登记机构和版权代理登记机构四级机构。具体设计如图 2-5 所示:

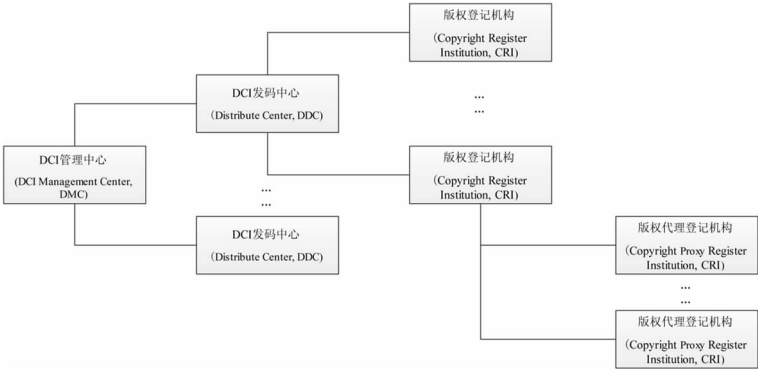


图 2-5 DCI 发码体系

版权登记机构下设版权代理登记机构,著作权人可以委托版权代理登记机构进行登记,为数字作品版权登记提供便利条件;顶级管理机构、多级 DDC 这种可扩展的设计让 DCI 发码体系不仅能解决中国所面临的版权问题,还能推动 DCI 走向世界,解决全球面临的数字版权问题。

2. 查重方案

在 DCI 发码过程中,涉及三种查重:第一种是登记机构对登记作品进行数据项查重,判断关键数据项是否重复;第二种是登记机构对合同备案中数字作品授权查重,判断合同中作品的权利转移是否合法;第三种是多 DCI 发码中心并行处理数字作品特征比对。第三种查重在实际业务操作过程中会面临多发码中心和海量数据的问题。

1)同步机制解决方案

每个发码中心维护所辖发码中心的服务器信息,避免查重时首先向管理中心请求的操作,同步机制实现过程如图 2-6 所示:

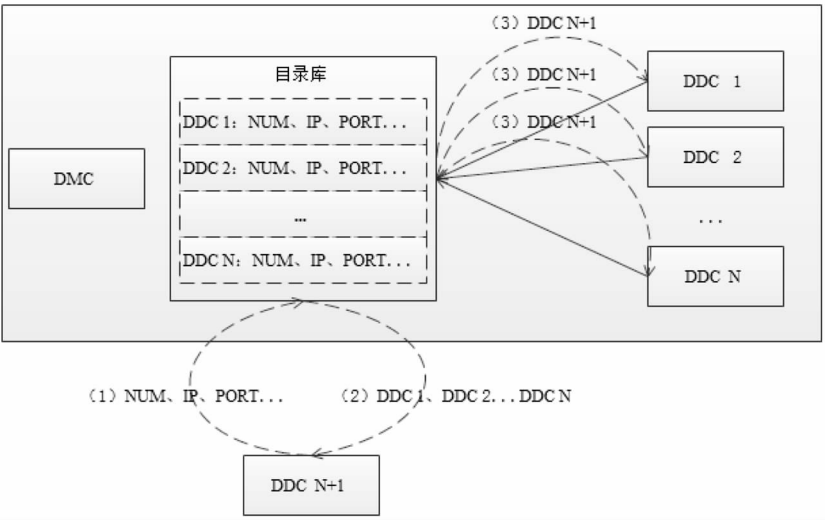


图 2-6 同步机制

2)数据库切分解决方案

如果把单点访问变成多点访问,就能降低处理的数据量,提高处理效率,这就用到了数据库切分,本文采用水平切分中的区域切分。数据库切分就需要解决全局唯一主键生成问题,生成策略如图 2-7 所示:

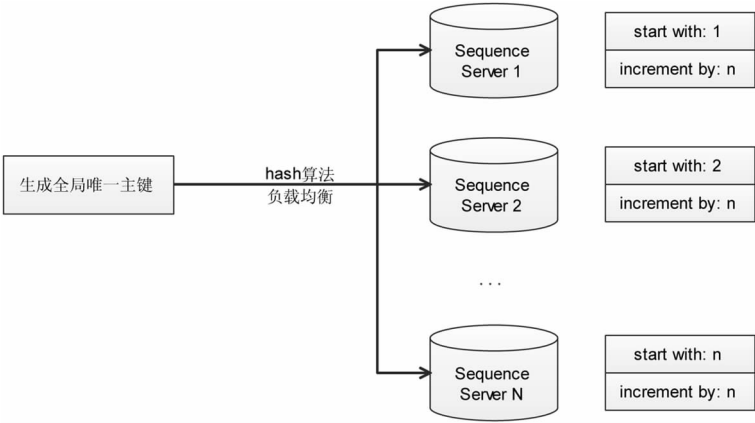


图 2-7 全局唯一主键生成策略

3. 解析方案

通过 DCI 查询作品基本信息、权利人和权属信息,在发生权利纠纷时为权利人自证提供依据,并

在数字作品交易过程中确认数字作品版权的真伪,为实现网络环境下实现数字作品版权的自动监测、跟踪奠定基础。DCI 解析原理如图 2-8 所示:

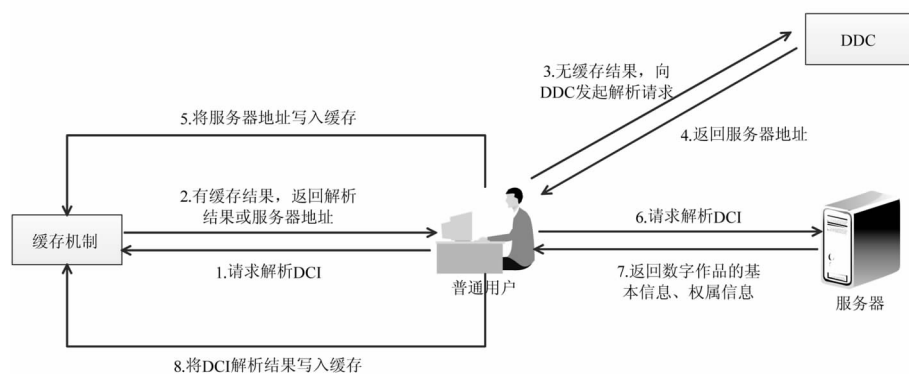


图 2-8 DCI 解析原理

2.3 核心系统设计与实现

为实现 DCI 发码与解析,需建设管理中心和发码中心两个系统,并实现发码中心系统与现有数字作品版权登记系统的对接。其中,管理中心系统主要实现 DDC 管理和服务器信息维护、同步等功能;发码中心系统主要实现查重、DCI 编码分发、版权查询、服务器信息管理等功能。通过管理中心系统、发码中心系统的建设以及与数字作品版权登记系统的对接,实现快速可靠的 DCI 发码方案和准确高效的 DCI 解析方案。

DCI 发码、解析是通过管理中心、发码中心和版权登记机构三个系统之间相互协作实现,需要对这三个系统的整合方案进行研究。整合方案如图 2-9 所示:

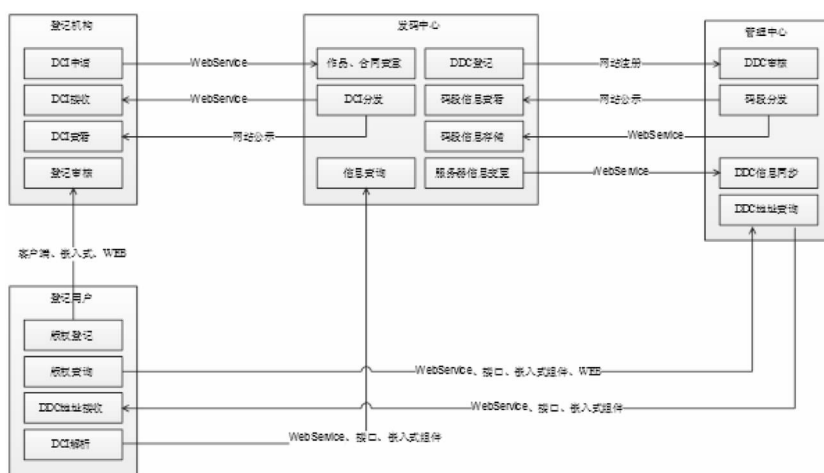


图 2-9 系统整合方案

2.4 总结与展望

本文设计的核心系统是国家科技支撑计划项目“基于版权的数字内容服务平台及应用示范(2012BAH04F00)”的部分内容。核心系统已完成与中国版权保护中心数字作品版权登记系统的对

接,解决了 DCI 发码问题,已投入使用,效果良好。

数字版权保护工作是一个系统工程,本文只对目前应用中最急迫、最核心的 DCI 发码和解析技术进行了研究。在数字版权业务处理过程的安全性、抗抵赖性等方面还需要深入的研究和探讨,这将需要突破更多的关键技术。

参 考 文 献

- [1] 全球文化产业发展简述 [EB/OL]. <http://artbank.people.com.cn/n/2012/0709/c209261-18477576.html>
- [2] 陈磊. DCI 体系,能否破解网络版权迷局 [N]. 2011-01-10
- [3] 杨树林,孙丽,胡洁萍. 数字版权保护中 DCI 编码方案及数字证书研究 [J]. 北京印刷学院学报, 2012, 20(6): 38-41
- [4] 方圆. DCI 体系有望助力版权维权机制创新 [N]. 2011-01-20
- [5] 杨东波,孙秀萍. 数字资源唯一标识符发展述略 [J]. 图书馆学研究, 2014, 04: 45-48
- [6] 赵长欣. 基于数字版权唯一标识符解析系统的设计与实现 [D]. 北方工业大学, 2012
- [7] ASL. DNS 全面解析 [J]. 电脑迷, 2014, 05: 72-73
- [8] 朱少强. 数字版权保护(DRM)技术标准的比较研究 [D]. 华东师范大学, 2004

The Research and Realization of Copyright Protection Core Architecture Based on DCI

Wang Shuai

(School of Computer Science and Technology, North China Univ. of Tech., Beijing, 100144)

Abstract This paper puts forward a copyright protection plan based on Digital Copyright Identifier (Digital Copyright Identifier, DCI), which includes three core contents: digital product copyright management based on DCI, it includes registration of digital product right, registration of contract authorization, extraction of digital content characteristic, building copyright database and characteristic database of digital product; the third party expense settlement and measurement based on DCI, sending back transaction information to copyright registration institution when digital product consumption platform sales a product with DCI, then obligee can access the institution to obtain the consumption records; infringement monitoring, it means obligee can apply to monitoring or report the infringement website, and the copyright management service institution obtains infringement evidence and generates a monitoring report.

The innovation of this paper is the prompt ion of the framework of product copyright protection based on DCI and designing solution of core framework by designing DCI code. In this paper, DCI coding system is designed to a distributed and extensible architecture; this architecture helps to DCI distribution and creates conditions for the promotion of DCI as an international standard.

Key words DCI; digital resource identifier; copyright protection; copyright service

时间序列数据挖掘 在 FY-3 MERSI 数据上的应用研究

研究生 田 圆

导师 曹丹阳

(北方工业大学计算机学院 北京 100144)

摘 要 气象局研究人员利用对地观测(Earth View, EV)和对冷空观测(Space View, SV)的数字量化输出值(DN 值)对风云三号(FY-3)搭载的中分辨率光谱成像仪(MERSI)进行绝对辐射定标,通过观察发现 SVDN 值受到一些因素的影响导致辐射定标结果不准确。本文旨在通过研究并改进时间序列数据挖掘的相关算法,将其应用于分析 FY-3 MERSI 在轨冷空观测值的影响因素,使得挖掘分析结果对辐射定标具有指导作用。本文的主要内容及创新点如下:

1. 本文根据课题中所要研究的时间序列数据时间维度高这一特点,提出了一种基于线段间夹角的时间序列分段线性表示方法。

2. 针对影响 SVDN 的特征维较多导致时间序列聚类运行速度慢、结果不精确这一问题,提出了一种基于主成分分析的欧氏距离时间序列聚类方法。该算法能够在保证聚类准确度不变或更高的前提下,通过降低特征维度来降低算法的运行时间。

3. 本文在上述两种算法的基础上设计并实现了 FY-3 MERSI 时间序列挖掘展示平台。

关 键 词 FY-3 MERSI; 辐射定标; 数据挖掘; 分段线性表示; 时间序列聚类

作者简介 田圆,女,1990 年出生,硕士研究生,主要研究方向为时间序列数据挖掘。

E-mail:448588273@qq.com

1 绪论

气象局研究人员利用对地观测(Earth View, EV)和对冷空观测(Space View, SV)的数字量化输出值(DN 值)对 FY-3 卫星上搭载的中分辨率光谱成像仪(MERSI)进行绝对定标^[1]。通过观察近几年 FY-3 在轨运行时 MERSI 获取到的 SV_{DN} 值,发现 SV_{DN} 的值受到一些因素的影响,影响了辐射定标的精度,最后则会导致遥感应用的不成功。

近些年来,随着各个领域中时间序列数据的日剧增加,越来越多的国内外学者更加重视时间序列数据挖掘。对时间序列进行挖掘具有非常重要的实际应用价值。目前将时间序列数据挖掘技术与气象卫星观测数据的分析结合在一起的课题研究比较少,本课题就是通过研究并改进时间序列挖掘相关算法后,将其应用于分析 FY-3 MERSI 在轨观测值的影响因素。

2 正文

2.1 气象相关知识

风云三号卫星(FY-3)是中国第二代极轨道气象卫星。中分辨率光谱成像仪(medium resolution

spectral imager, MERSI) 是风云三号 B 星上搭载的主要仪器之一。在本课题中, 我们利用辐射传输正演计算得到的大气顶反射率以及 MERSI 对地球观测和对冷空观测(作为辐射零点)的计数值来进行辐射定标, 确定各波段定标斜率。公式为:

$$RefFactor(i) = a(i) (EV(i) - SV(i)) \quad (2.1)$$

$$RefFactor(i) = Ref(i) (d_0/d)^2 \cos(SolZ) \quad (2.2)$$

其中, a 为定标斜率, i 为波段, Ref 为正演计算出的大气顶反射率, $RefFactor$ 为反射率因子(0~100%), $SolZ$ 为太阳天顶角, $(d_0/d)^2$ 为日地距离修正因子, EV 和 SV 分别为 MERSI 对地球和对冷空观测的计数值^[1]。其中 EV 的计数值 DN 我们可以通过其它各种手段测得, 但是 SV 的计数值 DN 我们只能通过卫星观测才能得到。从公式(2.1)可以看出, SV 的 DN 值正确与否影响着定标系数的精确度, 是极为重要的数值。

2.2 时间序列数据挖掘

时间序列是一种重要的高维数据类型^[2], 它是指按照时间次序排列且存在相互关联的一组随机序列, 本文中的 SV , 不同温度值以及不同角度值的数据均属于时间序列数据。近些年来, 越来越多的国内外学者开始重视时间序列数据挖掘。对时间序列进行挖掘具有非常重要的实际应用价值。

时间序列数据挖掘被广泛研究的方向有: 1) 时间序列的表示^[3]; 2) 时间序列预测; 3) 时间序列分类^[4]; 4) 时间序列聚类^[5]; 5) 时间序列异常点检测; 6) 时间序列数据可视化^{[6][7][8]}。根据本课题的研究目的, 本文重点研究时间序列表示中的分段线性表示方法以及时间序列聚类的相关算法。

2.3 时间序列分段线性表示

目前已有的分段线性表示算法都存在其各自的优缺点, PAA 表示方法计算简单, 但是拟合比较粗糙, 对时间序列的特征值如极大值点、极小值点、拐点等无法准确的拟合; PLR_PF 方法的拟合效果与 PAA 表示方法相比, 有了明显的提高。但只考虑了时间序列中的局部极值点, 无法选出非局部极值点但是时间序列趋势转折重要的点; SEEP 方法对于斜率变化范围明显的时间序列段, 有着很好的拟合效果。相较前两种方法, SEEP 大大降低了拟合误差, 但是针对波动范围小而斜率变化范围大的时间序列, SEEP 方法无法准确的选取分段点从而导致压缩率低的问题。

2.4 时间序列聚类

目前已有的时间序列聚类方法都存在其各自的优缺点, 基于原始数据的时间序列聚类方法比较简单, 可以直接对时间序列进行聚类分析, 但当时间序列的特征维数过高并且含有大量噪声的时候, 该方法的效率会明显下降, 有时甚至得出错误的结论。基于特征的时间序列聚类方法虽然可以通过对时间序列降维提高算法效率, 但是降维后有时会丢失原始数据中的重要信息。基于模型的时间序列聚类的优点在于可以处理含有噪声、不确定性、程度变化的时间序列等, 但其缺点在于过程太过复杂, 需要通过对时间序列建立模型, 使用模型的参数、系数等信息进行挖掘聚类。

3 基于线段间夹角的时间序列分段线性表示方法

本课题研究的 space view (SV_{DN}) 值、不同的温度值、以及不同的角度值的时间序列数据的最大特点在于其时间维度很高。故首先要对时间维度进行降维处理。降维方法我们选用 2.3 节中提到的分段线性表示方法。通过观察发现本文研究的这些时间序列数据中存在很多波动范围小但斜率变化范围大的区段, 若采用 2.3 中陈述的已有的分段线性表示方法无法很好的选择原始时间序列的分段点。故在本文中提出一种基于夹角的时间序列分段线性表示方法对本课题中的时间序列进行时间降维。

3.1 算法描述

目前在已有的时间序列分段线性表示算法中,基于斜率的时间序列分段线性表示算法 SEEP 无论在拟合误差和压缩率方面,还是在算法的时间复杂度上都有明显的优势。但是 SEEP 对于波动范围小但斜率变化范围大的时间序列(下图 3.1 加粗线表示)无法准确的识别分段点。通过对时间序列曲线的直观观察,发现当线段间的夹角越小的时候,曲线的变化越明显。这时导致线段形成夹角度数小的点就成为了重要的点,再通过计算相邻的两个夹角顶点之间的垂直距离(下图 3.1 大括号表示),排除重要点中波动范围小的点,即从重要点中选出时间序列的分段点。本文依据该思想提出了基于线段间夹角的分段线性表示方法 (PLR_AG)。

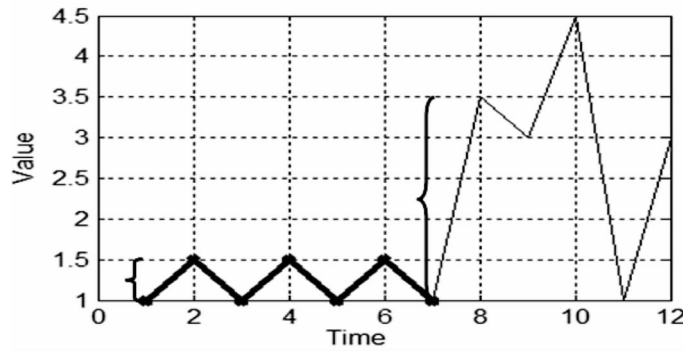


图 3.1 波动范围小但斜率变化范围大的时间序列

该方法首先计算连接同一点的两条相邻线段间的夹角,通过判断该夹角的度数与用户所给的角度参数阈值的关系来判断所形成夹角的顶点是否为重要点,若该点被确定为重要点,那么再计算该顶点与其前一个序列点之间的垂直距离,若垂直距离大于用户所给的距离约束阈值,则该重要点被选为分段点。最后依次将选出的分段点用直线段相连接,就得到了该方法。

本文提出的时间序列分段线性表示方法关键在于计算夹角的大小,为方便起见,本文将形成夹角的第一条线段称为 l_1 ,第二条线段称为 l_2 ,两条线段间的夹角称为 α , l_1 与 x 轴的夹角为 α_1 , l_2 与 x 轴的夹角为 α_2 。通过总结发现当 α_1 和 α_2 类型一致(即同为锐角或同为钝角)时,无论 α 是何种类型,其计算公式均为 $\alpha = \pi - |\alpha_1 - \alpha_2|$;当 α_1 和 α_2 的类型不一致(即一个为锐角且另一个为钝角)时,无论 α 是何种类型,其计算公式均为 $\alpha = |\alpha_1 - \alpha_2|$ 。公式 3.1 即为夹角 α 的计算公式:

$$\alpha = \begin{cases} \pi - |\alpha_1 - \alpha_2|, & \alpha_1 \alpha_2 \geq 0 \\ |\alpha_1 - \alpha_2|, & \alpha_1 \alpha_2 < 0 \end{cases} \quad (3.1)$$

3.2 算法实现

算法名称:基于线段间夹角的时间序列表示算法 PLR_AG

算法输入: X , $|X| = n$, 时间序列; β , 夹角阈值; d , 垂直距离阈值;

算法输出: X_{new} , $|X_{new}| = n$, 拟合后的新时间序列;

具体步骤:

(1) $i=0; j=1; k=2$;

(2) $X_{PLR} = \{x_1 = (v_1, 1)\}$; //时间序列的第一个点为分段点

(3) for (count=1; count<n-1; count++)

(4) $\alpha_1 = \arctan((x[j] - x[i]) / (j - i))$; //计算 l_1 与 x 轴的夹角

```

 $\alpha_2 = \arctan((x[k] - x[j]) / (k - j));$  // 计算  $l_2$  与  $x$  轴的夹角
if ( $\alpha_1 \leq 90 \wedge \alpha_2 \leq 90 \mid \alpha_1 > 90 \wedge \alpha_2 > 90$ ) // 两条线段的夹角同为锐角或同为钝角
then  $\alpha = \pi - \text{abs}(\alpha_1 - \alpha_2);$ 
(5) else // 两条线段的夹角类型不一致时
 $\alpha = \text{abs}(\alpha_1 - \alpha_2);$ 
(6) if ( $\alpha < \beta$ ) // 若夹角小于所给参数  $\beta$ , 即导致其产生该夹角的点为重要点
then  $dis = x[j] - x[i];$  // 若该点是重要点, 计算该点与上一个点的垂直距离
if ( $dis \geq d$ ) // 若计算出的垂直距离大于用户给定的距离约束条件
then  $X_{PLR} = X_{PLR} + \{x_{count} = (v_{count}, count)\};$ 
 $i = count; j = count + 1; k = count + 2;$ 
 $X_{PLR} = X_{PLR} + \{x_n = (v_n, n)\};$  // 时间序列的最后一个点为分段点
(8) output  $X_{new} = \{x'_1, x'_2, \dots, x'_i, \dots, x'_n\};$  // 得到新时间序列
实验数据及结果分析

```

在实验中, 采用了 Keogh Eammon^[9] 等人提供 8 条时间序列, 这 8 条实验数据均取自不同领域(见表 3.1), 来比较算法的性能, 数据能够具有广泛性。

表 3.1 数据集介绍

序列名称	序列长度
Burst	9382
Chaotic	1800
Earthquake	4097
Leleccum	4320
Ocean	4096
Powerplant	2400
Speech	1021
Tide	8746

(1) 4 种 PLR 算法的性能比较:

由于时间序列分段线性表示的压缩率与拟合误差密切相关, 所以对不同算法均采用相同的压缩率, 取压缩率为 80%。实验结果如表 3.2 所示:

表 3.2 相同压缩率下不同算法的拟合误差比较

算法数据集	<i>PAA</i> 表示	<i>PLR_{PF}</i> 表示	<i>SEEP</i> 表示	<i>PLR_{AG}</i> 表示
Burst	1.21	0.89	0.07	0.08
Chaotic	2.33	1.63	0.49	0.42
Earthquake	3.25	2.11	3.50	2.25
Leleccum	1.67	0.88	0.91	0.87
Ocean	1.27	0.38	0.51	0.48
Powerplant	4.13	2.23	0.59	0.48
Speech	2.87	1.52	1.72	1.61
Tide	3.12	2.29	3.32	2.46

在表 3.2 中,黑体字加粗表示的为拟合误差最小的算法。从表中可以看出,在不同领域的时间序列数据中,PLR_AG 算法的拟合误差始终保持相对较小的水平,拟合误差平均降低约 15%,说明 PLR_AG 算法在时间序列的分段线性表示方面能够通过降低拟合误差来取得较好的效果。

(2)比较同一条时间序列在压缩率分别等于 60%,70%,80%,90%下不同方法的拟合误差,实验结果如图 3.2 所示:

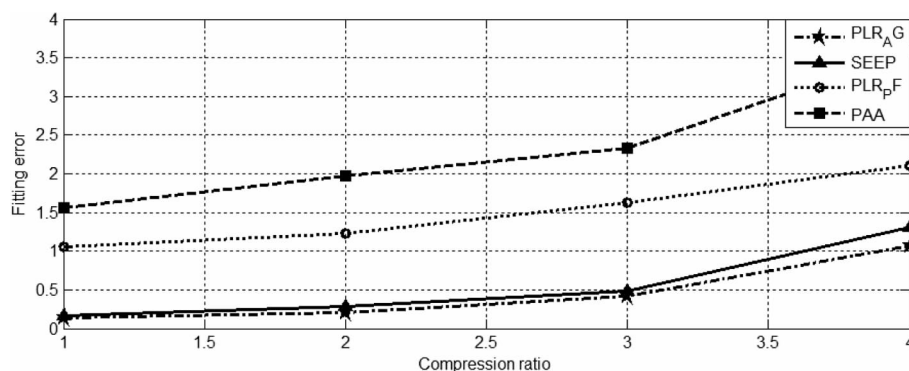


图 3.2 不同压缩率下 4 种分段方法拟合误差的曲线

从图 3.2 中可以看出时间序列的压缩率与拟合误差之间的关系,压缩率越高,拟合误差也随之增高。通过观察图发现在不同的压缩率下 PLR_AG 算法的拟合误差始终是最小的,该算法较 PAA 算法及 PLR_PF 算法在拟合误差上有明显的提高,较 SEEP 算法也有小幅度的提高。

4 基于主成分分析的欧式距离聚类方法

为了挖掘这 7 个温度以及 7 个角度对每个通道 SV_{DN} 值的影响关系,结合时间序列数据挖掘的思想,即对这些不同的时间序列进行聚类分析,聚类结果能够说明哪些温度值以及哪些角度值对不同通道的 SV_{DN} 值是有影响的。由于本课题中不同的影响因素之间存在一定的相关关系,可以通过降低特征维度后再对其进行时间序列聚类分析。本文结合基于原始数据和基于特征的时间序列聚类方法的优点提出了基于主成分分析的欧式距离时间序列聚类方法。

算法实现

算法输入: $\{x_1(t), x_2(t), \dots, x_m(t)\}, t=1, 2, \dots, T$, 时间序列数据集; threshold, 欧氏距离判定阈值

算法输出:

$result = \{R1 \cup R2 \cup \dots \cup Rn\}, Rn = \{x_i(t) \cup x_j(t) \dots x_{kaa}(t)\} (i, j, k=1, 2, \dots, m; t=1, 2, \dots, T)$, 时间序列聚类结果

步骤 1 对时间序列数据集进行标准化;

$m = \text{mean}(X_{maa}(t), axis=0)$

$data = m$

步骤 2 协方差矩阵的求解;

$C_m = \text{cov}(\text{transpose}(data))$

步骤 3 计算特征值与特征向量(系数矩阵);

$\text{eigvalue}, \text{eigvectors}_n \times t = \text{linalg. eig}(C)$

步骤 4 对系数矩阵求欧氏距离;

$dis = \text{pdist}(\text{eigvectors}, 'euclidean')$

步骤 5 根据欧氏距离公式得到欧式距离矩阵;

$dis = \{d_{ij}\}, i=1, 2, \dots, n, j=i+1$

步骤 6 根据求得的矩阵以及输入的阈值 threshold 得出最终聚类结果; $result = \{R1 \cup R2 \cup \dots \cup Rn\}, Rn = \{x_i(t) \cup x_j(t) \dots x_k(t)\} (i, j, k=1, 2, \dots, m; t=1, 2, \dots, T)$

实验结果及分析

随着时间序列数据集的维度不断增加, 比较两个算法的 Cophenetic 系数(该系数越接近与 1, 说明聚类效果越好)和平均运行时间, 结果如图 4.1 与图 4.2 所示: 其中“□”代表直接使用欧氏距离的时间序列聚类方法, “*”表示基于 PCA 的欧氏距离时间序列聚类方法。

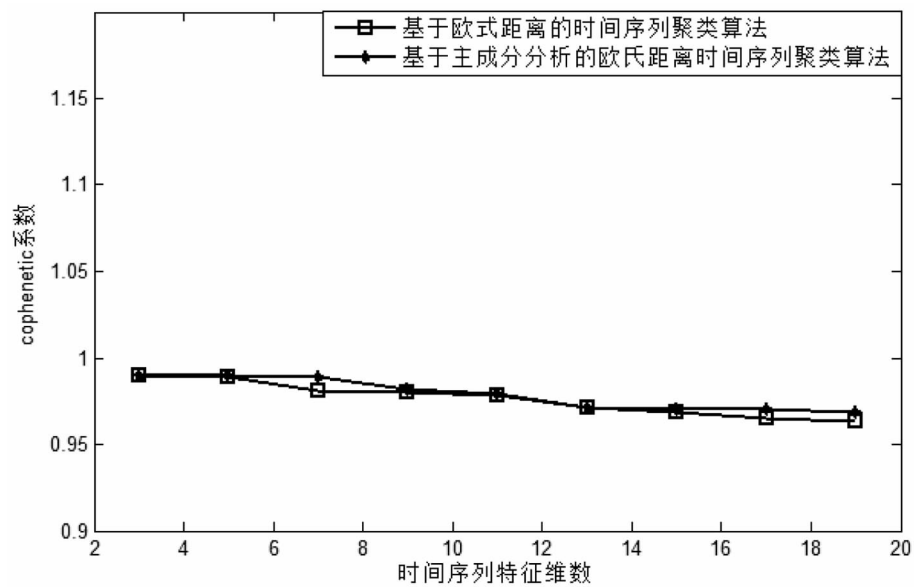


图 4.1 算法 Cophenetic 系数比较

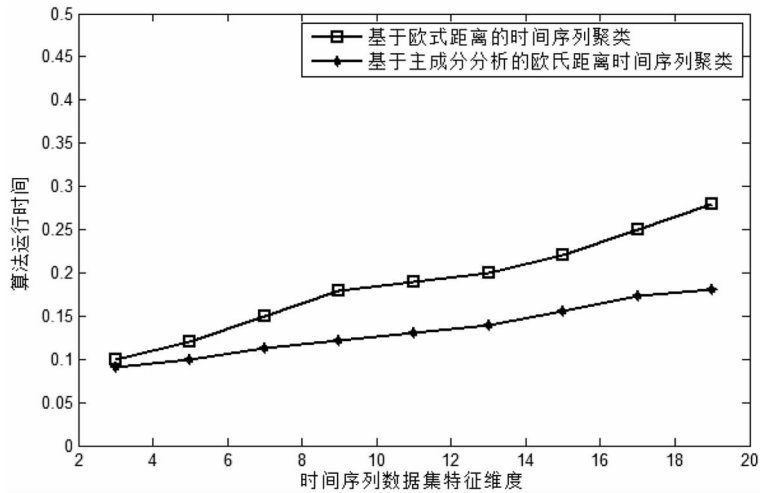


图 4.2 算法运行时间比较

结合图 4.1 与图 4.2 可以看出,本文中的算法能够有效的保证算法的准确性,在 Cophenetic 系数不降低的前提下,基于 PCA 的欧氏距离度量算法的运行时间要明显优于基于传统欧式距离的时间序列聚类算法的运行时间。说明本文中提出的算法在时间序列数据集的聚类分析上能够通过降低维度,计算系数矩阵的欧式距离来取得较好的效果。

5 FY-3 MERSI 时间序列挖掘展示平台

5.1 FY-3 MERSI 时间序列挖掘平台简介

该平台由三大模块构成,分别是时间序列分段线性表示模块、时间序列聚类分析模块以及 FY-3 MERSI 时间序列数据应用展示模块。其中前两大模块的功能分别是对前两章提出的两个算法的效果展示与算法之间的比较,通过前两个模块的功能可以验证本文提出的两个算法的优势。第三个模块的功能是对本课题进行分步分析挖掘的效果展示,通过模块三可以得出本课题的最终结论。平台如图 5.1 所示:



图 5.1 平台展示图

5.2 FY-3 MERSI 序列时间降维预处理

通过本文中提出的基于夹角的分段线性表示方法,将已有的 20 个通道的 SV_{DN} 序列,7 个温度序列,7 个角度序列均从 365 个时间维度降低到 120,为下一步使用降维后的时间序列数据集分组进行聚类分析做准备。结果如图 5.2 至 5.4 所示。

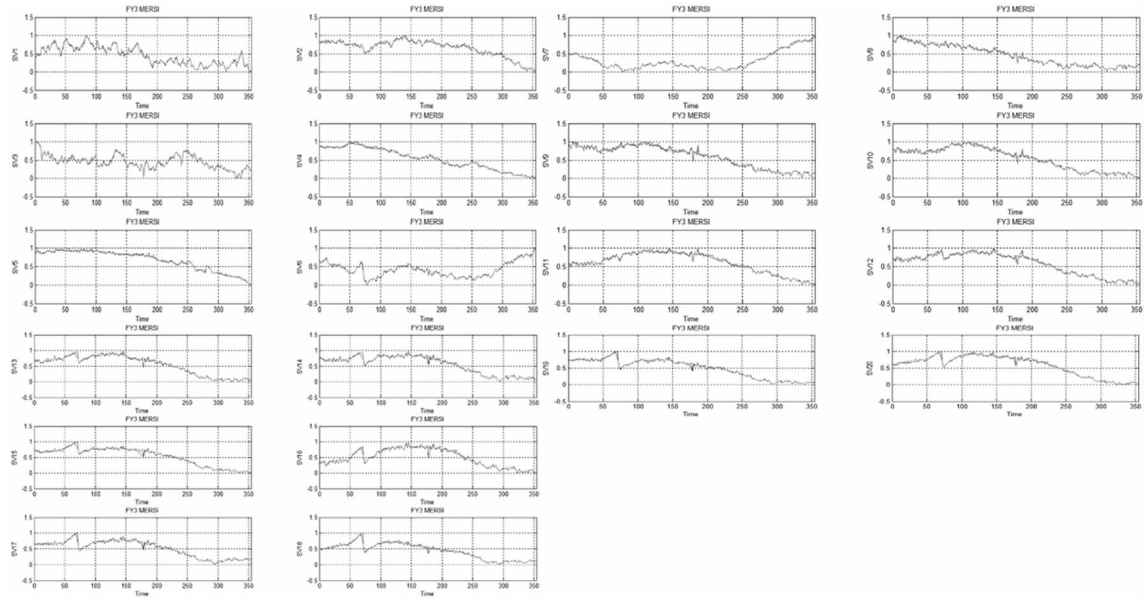


图 5.2 SV_{DN} 值降维后效果图

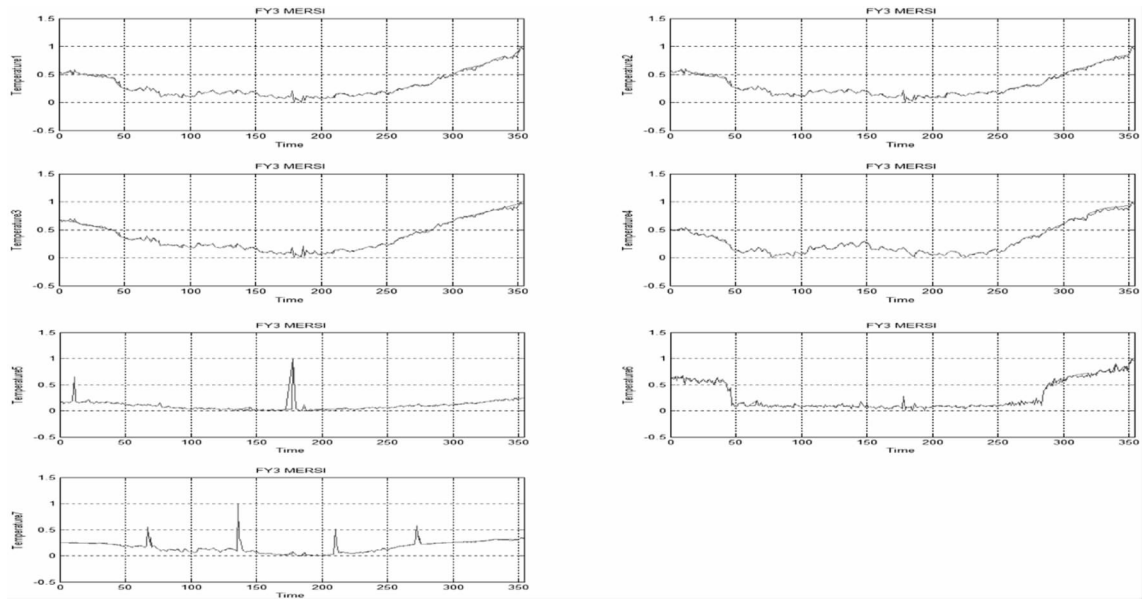


图 5.3 温度值降维效果图

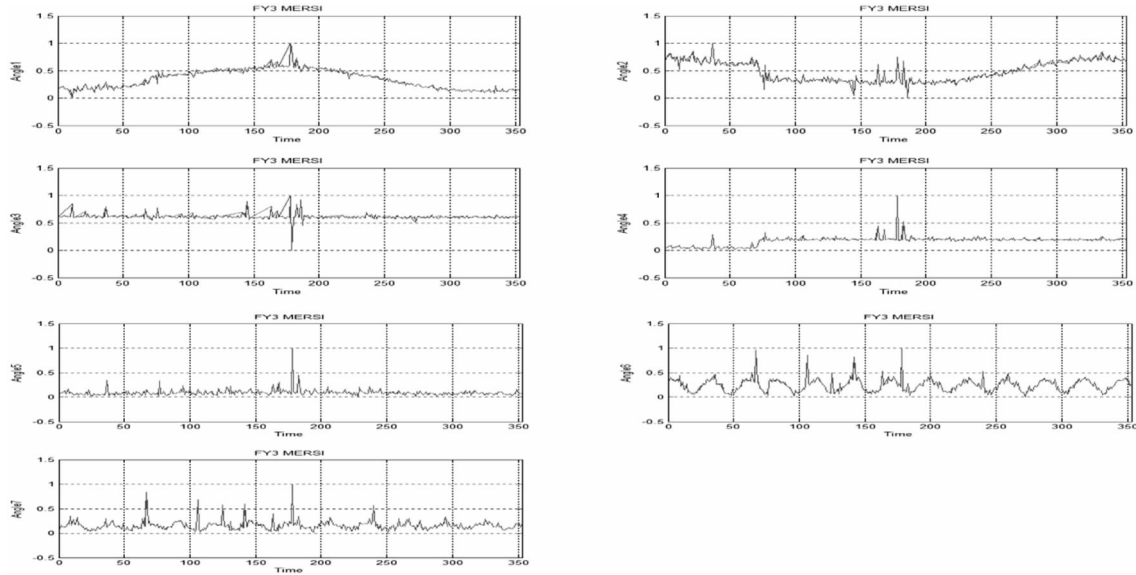
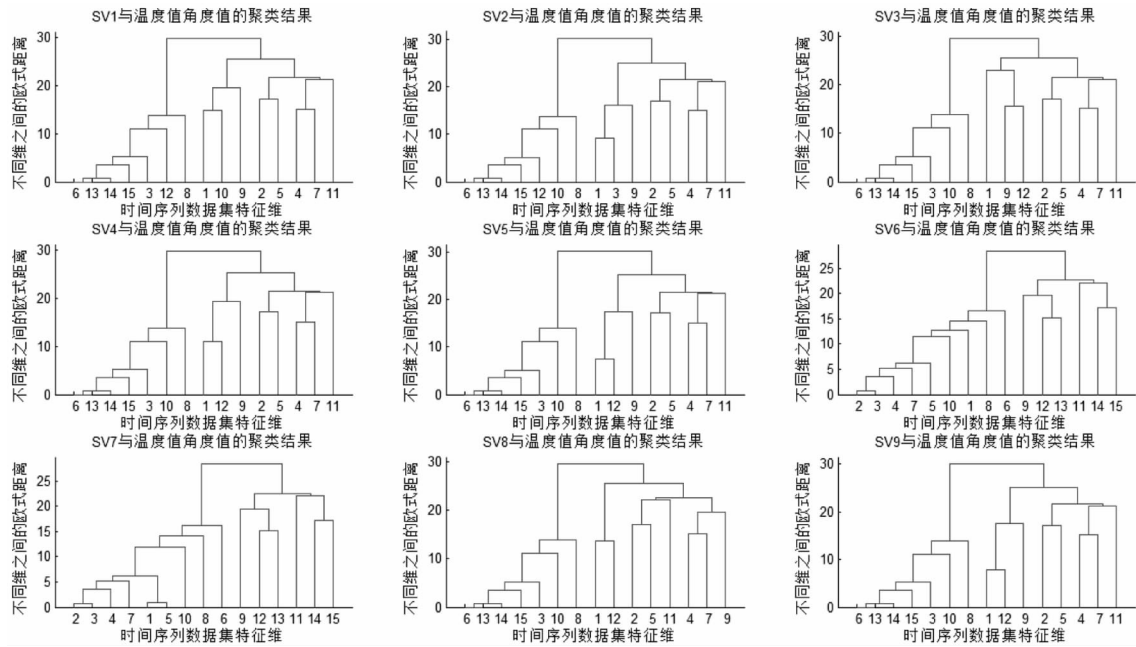
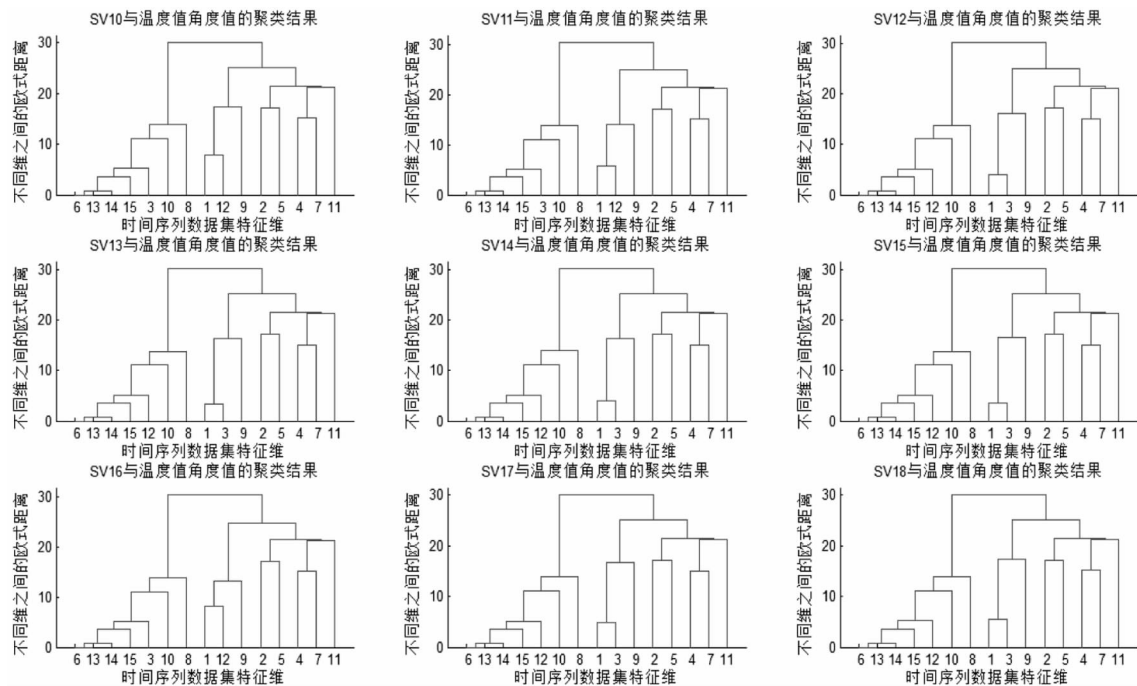
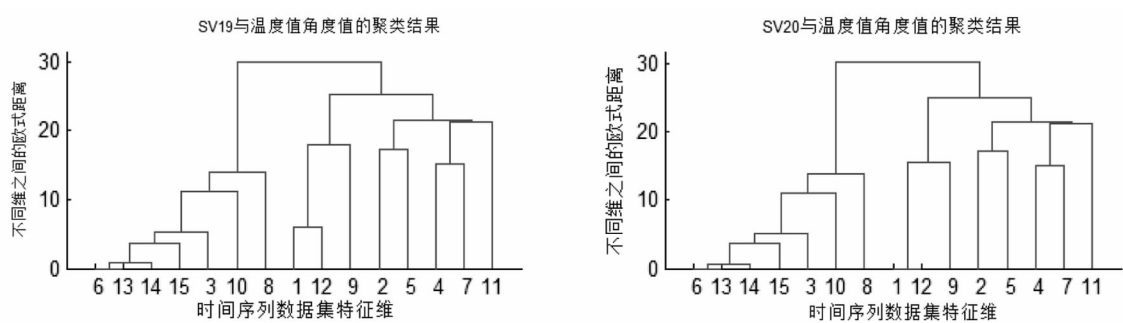


图 5.4 角度值降维后效果图

5.3 FY-3 MERIS 特征降维并聚类

对降维后的不同时间序列分组进行聚类分析,结果如图 5.5 至 5.7 所示。

图 5.5 通道 1 至 9 的 SV_{DN} 值与温度值角度值聚类结果

图 5.6 通道 10 至 18 的 SV_{DN} 值与温度值角度值聚类结果图 5.7 通道 19,20 的 SV_{DN} 值与温度值角度值聚类结果

5.4 本课题结论分析

通过以上的时间序列数据处理及聚类,根据聚类结果计算同一类中不同时间序列的相关系数,可以得出本课题的结论:1)在 7 个温度值中,主要是 Cool1,BB,Bracket1,Bracket2,KMirror 对 SV_{DN} 值有影响,且为负影响;2)在 7 个角度值中,主要是 EVCAzi,EVCZen,EVCLat 对 SV_{DN} 值有影响,其中 EVCAzi 主要为正影响,EVCZen,EVCLat 为负影响;3)通过分析发现 6,7 通道;9,10,11 通道;2,12,13,14,15,17,18 通道;19,20 通道彼此之间的特性很相似,再更进一步的工作中可以将它们划分在一起分析挖掘。

参 考 文 献

- [1] 孙凌,郭茂华等.基于敦煌场地定标的 FY-3 MERSI 反射太阳波段在轨响应变化分析[J].光谱学与光谱分析.

2012

- [2] Keogh E, Kasetty S. On the need for time series data mining benchmarks: a survey and empirical demonstration [J]. Data Mining and Knowledge [J]. 2012.
- [3] 原继东, 王志海. 时间序列的表示与分类算法综述[J]. 计算机科学. 2015. 03
- [4] 贾澎涛, 何华灿. 时间序列数据挖掘综述[J]. 计算机应用研究. 2007
- [5] 韩忠明, 陈妮等. 面向热点话题时间序列的有效聚类算法研究[J]. 软件学报. 2012
- [6] YU J, HUNTER J, REITER E, et al. Recognizing visual patterns to communicate gas turbine time-series data [C]. Application and Innovations in Intelligent Systems. London: Springer, 2002
- [7] HOCHHEISER H, SHNEIDERMAN B. Visual queries for finding patterns in time series data[R]. Maryland: University of Maryland. 2002
- [8] HOCHHEISER H. Interactive graphical querying of time series and linear sequence data sets [D]. Maryland: University of Maryland. 2003
- [9] Keogh E, Folias T. The UCR Time Series Data Mining Archive [EB/ OL]. <http://www.cs.ucr.edu/~eamonn/TSDMA/index.html>. Irvine, CA: University of California, Department of Information and Computer Science, 2002

Time Series Data Mining in FY—3 MERSI Research

Tian Yuan

(School of Computer Science and Technology, North China Univ. of Tech., Beijing, 100144)

Abstract The Meteorological administration researchers use digital number (DN) of Earth View (EVDN) observation and Space View (SVDN) observation of medium resolution spectral imager (MERSI) which is a sensor carried on the FY—3 Meteorological satellite to do radiometric calibration. Through the observation of SVDN value, the researchers found the SVDN value was influenced by some factors, so the inaccurate SVDN value will lead to inaccurate radiation calibration result. The purpose of this paper is by means of research and improvement of relevant time series data mining algorithms, and then applied these algorithms to the analysis of SVDN value of FY—3 MERSI. In this paper, we use the nearly three year's data to find out which factors influence the SVDN value. The researchers of Meteorological administration can use the result of this time series data mining to guide they work. In this paper, the main content and innovation points are as follows:

1. Due to the characteristic of high time dimension of time series which will be researched in this paper, this paper proposes a method of piecewise linear representation of time series based on angle between lines (PLR_AG).
2. According to the high dimensions of feature that will lead to slow speed and wrong time series clustering result. This paper proposes the idea of time series clustering analysis method based on principal component analysis.
3. On the basis of the above two algorithm, a FY—3 MERSI time series mining platform was be

designed and implemented.

Key words FY-3 MERSI; radiometric calibration; data mining; time series piecewise linear representation; time series clustering

基于立体视觉的三维重建算法研究

研究生 孔德星

导师 范涵奇

(北方工业大学计算机学院 北京 100144)

摘 要 三维重建是计算机视觉和计算机图形学中重要的研究内容,是一个国家计算机科技发展进步的重要标志之一,在科研和工业上都有重要的用途。

三维重建最核心的内容就是准确地获取场景的深度信息。针对当前获取深度信息技术的种种缺陷,本文提出了一种从含有噪声和空洞的低分辨率深度图重建出高质量、高分辨率深度图的算法。算法基于以下假设:高分辨率深度图是由一组过完备的字典的稀疏线性组合生成而得到的,低分辨率深度图是从高分辨率深度图中采样而得。

本文算法分为三个步骤。第一步,将彩色相机和深度相机进行立体标定,这样可以将低分辨率深度图投影到高分辨率彩图上得到稀疏的高分辨率深度图。为了使深度相机标定过程自动化,本文提出了深度图中矩形平板的自动提取算法。第二步,本文在现有的图像分割方法上改进,同时利用稀疏的深度信息和高分辨率彩图信息,找出场景中深度连续变化的各个区域,将高分辨率深度图准确地进行分割;第三步,对于分割后彩图的每个区域,在贝叶斯框架下求解最优的高维稀疏向量,继而得到高分辨率深度图,最后重建出三维点云模型。

本文对当前流行的主要算法和本文算法进行了大量的实验对比。在 Middlebury 测试集上的实验结果的量化指标的综合对比和视觉效果都表明本文算法要优于其它算法。真实场景深度图的重建结果的对比中,本文算法的优点体现的更明显。

关 键 词 三维重建; 摄像机标定; 图像分割; 稀疏表示

作者简介 孔德星,男,1989 年出生,硕士研究生,主要研究方向为计算机图形学。

E-mail: dxkung@163.com

1 绪论

三维重建是指利用二维图像恢复物体或场景三维信息的一种计算机技术,是计算机视觉和计算机图形学中非常重要的研究领域之一,在国家和社会发展中承担着不可估量的作用。由于传统的成像技术是将现实的三维世界透视投影到二维图像上,损失了原有的深度信息,所以三维重建的核心技术就是准确地获取物体或场景的深度信息。目前获深度信息的工具和方法大致可分为三种:3D 激光扫描仪,立体匹配的方法和 ToF、Kinect 等深度相机。3D 激光扫描仪具有高精度,高密度等特性,可获得非常准确的深度图,但它扫描的速度太慢,并且价格非常昂贵。立体匹配^[1]是通过双目摄像机的立体标定,用一定的匹配算法寻找匹配点,然后计算视差(逆深度),进而得到深度信息。立体匹配比较廉价,但是获得的深度信息精度太低,匹配速度也很慢,容易受图像纹理的影响。ToF 深度相机和 Kinect 深度相机比立体匹配最大的优势是,它们可以不受场景纹理的影响快速获取深度信息,其准确度在 3D 激光扫描仪和立体匹配之间。它们在社会上已经有了非常广泛的应用。然而,这类设备获得的深度图

的分辨率都远远低于普通彩色相机的分辨率。另外,这些设备捕获的深度值含有很强的噪声,且常常含有空洞。

本文先用廉价快速的 Kinect 深度相机获取一张分辨率为 640×480 的低分辨率深度图。为了获取更高分辨率的深度图,本文添加了一个分辨率为 1920×1080 的普通彩色相机。彩色相机和深度相机进行标定后,根据双目立体视觉的相关理论,低分辨率深度图就可以投影到高分辨率的彩图上。由于深度图的分辨率远远低于彩色图片,深度图投影到彩色图片后,只能得到非常稀疏的深度图,甚至出现成片区域缺少深度值的情形。这个客观事实可以理解为从高分辨率的场景深度图中随机观测到一部分深度值,并将这些深度值赋予对应的彩图像素。由于观测是随机的,观测的过程又不够理想,因此这些深度值含有大量的空洞和噪声。受信号多形态分析^[2]和压缩感知理论^[3]的启发,本文使用一组过完备基的稀疏线性组合^[4]来表达高分辨率深度图,并将重建问题归结为高维稀疏信号的求解问题。求解时,将已观测到的深度值作为约束条件,完全不使用彩图纹理信息,从而摆脱了彩图纹理的影响。Torralba 等^[5]发现,彩图纹理中的边界往往是深度变化跳跃的地方,同一物体的深度通常是连续变化的。本文从彩图纹理信息出发,结合已观测到的深度值,找出深度变化连续的区域,对彩图进行分割,然后对各个区域分别进行重建,所得的深度图更为准确。

2 相关技术研究

在捕获场景的低分辨率深度图后,通过增加高分辨率彩色图像来获得高分辨率深度图的想法并不是什么全新的方法,学术界已经进行了大量的研究。现有的方法大致可以分为两种,基于滤波器的和基于最优化理论的。

基于滤波器的算法是最为常见的,如 JBU(joint bilateral upsampling)^[6]、NAFDU(noise-aware filter for depth upsampling)^[7]和 JABDU(joint-adaptive bilateral depth map upsampling)^[8]。JBU 是一种非常受欢迎的深度图上采样算法。该算法思想简洁明了,甚至能做到实时^[9],但是非常容易受到彩图纹理的影响,产生纹理复制的问题。NAFDU 进一步考虑了深度图中的噪声特性,提出了混合权重的方法,尽可能地降低 JBU 算法中彩图纹理对深度图的不良影响,提高了重建的质量。Kim 等^[8]提出的 JABDU 算法是在双线性插值算法上采样得到的深度图的基础上,利用彩图纹理信息对该深度图进行细化,解决双线性插值导致的图像轮廓模糊问题。这些基于滤波器的上采样算法的优点是简单直观,时间复杂度低,产生的深度图比较光滑,但是其缺点也非常突出,容易平滑深度图中物体的边界。

作为基于优化的算法,为了有效利用彩图信息,Diebel 等^[10]构造了一个双层的马尔科夫随机场,节点之间的权重是彩图像素的相似度的度量。在此基础上,Park 等提出了 HQDMU(high quality depth map upsampling for 3d-tof cameras)算法^[11]。该算法在边界处考虑多种影响因素来确定节点之间的权重,其他节点之间的权重由彩图纹理局部区域的相似性来确定。为了处理边界问题,Ferstl 等^[12]使用彩图纹理的二阶 TGV(total generative variation)达到分片的目的,通过最优化一个凸优化问题达到上采样的目的。基于优化的算法,其结果要好于基于滤波器的上采样算法,但仍然有很严重的问题:HQDMU^[11]和 ATGV^[12]在上采样过程中容易受到彩图纹理的影响,纹理复制的问题依然存在,而且它们均假定低分辨率深度图不含空洞,投影到彩图上也是均匀的。

3 彩色相机和深度相机的立体标定

本文采用廉价的 Kinect 深度相机,其分辨率为 640×480 ,为了得到更高分辨率的深度图,本文使

用了一台分辨率为 1920×1080 的普通彩色相机。由于两种相机得到的是不同视点下的图像,因此需要将这两个相机进行标定。

深度相机测量采用的是不可见的红外光,而彩色相机工作在可见光的范围内。Herrera 等人^[13]巧妙地将标定彩色相机所用的棋盘格紧密地贴在矩形平板上,然后利用棋盘格与矩形平板共面的特点,将两个相机标定在一起。Raposo 等人^[14]改进了该算法,提高了其精度。然而,这两个算法进行标定时,需要使用者在几十幅深度图中用鼠标勾勒出矩形的轮廓。本文将彩色相机和深度相机的视点调整到大致相同,使得用于标定的矩形平板在高分辨率的彩图和低分辨率的深度图中所处的相对位置比较接近。利用这个特点,本文从彩图中的棋盘格格点的位置获得深度图中矩形平板的大致位置,继而从深度图中自动提取出矩形平板,摆脱了人工交互勾勒的过程,完善了 Raposo 等人^[14]的标定算法。提取矩形平面的具体算法如算法 3.1。

算法 3.1: 矩形平面提取算法

Require: 彩图 I^c , 深度图 D^d

Ensure: 彩色相机与深度相机视点相近

1: 从 I^c 中获得棋盘格的格点 p_c ;

2: 依据 I^c 、 D^d 的分辨率,将 p_c 线性映射到 D^d ,得到 p_e ;

3: 从去除了畸变后的深度图中,获得 p_e 的深度值 d^d ,得到三维点 z_e ;

4: 由 Z_e ,通过 PCA (Principal Component Analysis) 算法获得平面的法向 n_e ;

5: 采用 Least Median of Squares 算法^[15],找出属于矩形平板的所有三维点。

4 基于区域合并算法的场景分割

场景中物体的深度通常都是不同的,不同物体之间的边界的深度值往往也是不连续的。在这些深度不连续的地方,如果不加区分地进行重建,得到的结果的误差往往是非常大的。为了解决该问题,本文首先通过联合使用彩图纹理信息和投影到彩图上的稀疏深度信息来找出物体之间的边界,将场景分割成多个不同区域,然后对这些不同区域分片重建深度图,从而减小重建的误差。

为了找到这些不同的区域,依据 Arbeleaz^[16]的建议,本文首先在彩图的 PB 边界^[17]上应用 watershed 算法,将彩图分割成众多的细小区域。为了确保绝大多数的区域没有跨越物体之间的边界,因此本文对彩图进行过分割。该分割结果通常包含了大约 3000 个区域。区域的合并问题实际上是一个区域相似度的度量问题,也就是 $Pr(r_i \neq r_j | D_{ij})$,其中 r_i 是区域 i 的标号, D_{ij} 是区域 i 和 j 的相似性的度量。Hoiem 等人提出的遮挡边界恢复算法^[18]综合考虑了区域的颜色和位置,边界的长短和光滑性,以及物体的朝向和深度,通过建立条件随机场(Conditional Random Field),从样本库中学习得到 D_{ij} 。

本文对 Hoiem 等人的算法^[18]做出了以下改进:1) 使用投影得到的稀疏深度值来计算区域深度,而不是从彩图中推测各个区域的深度;2) 纹理相似的区域都参与合并,而不仅限于相邻区域。虽然这些稀疏的深度值含有空洞和噪声,但仍然比推测得到的深度更为准确。利用这些深度值,该算法能够很好地将彩色纹理相近的区域区分开来,避免了对深度差别太大的区域进行合并。

5 深度图重建

5.1 重建算法

获得深度连续变化的区域后,本文分别重建每个区域的深度图。重建算法不使用任何彩图纹理信息,从而摆脱了纹理对重建结果的影响。如不特殊说明,本文下面叙述的高分辨率深度图是指分片得到的每个区域。

本文将高分辨率深度图记为 $x \in \mathbb{R}^n$, n 为像素的个数。 x 可以由 l 个 n 维的原子线性组合而得:

$$x = \Phi\alpha + \epsilon, \epsilon \in N(0, \sigma^2 I) \quad (5.1)$$

其中, $\Phi \in \mathbb{R}^{n \times l}$ 通常被称为字典,它的每个列向量称为原子。 $\alpha \in \mathbb{R}^l$ 为对应的系数向量,而 ϵ 为噪声。

字典 Φ 的表达能是至关重要的,它直接影响重建得到的深度图的质量。为了获得更强的表达能力,可以将多个不同的字典合并成为字典 Φ 。本文用于表达高分辨率深度图的字典含有四个不同类型的变换:Daubechies 小波基非常适合表达细小的局部特征;LDCT(Local Discrete Cosine Transform)能够很好地表达重复性的纹理;Grouplets^[19]能够很好的捕捉方向性、结构性的特征;Curvelet^[20]适合表达分片光滑的深度图。本文首先将高分辨率深度图展成一维向量 $x \in \mathbb{R}^n$ 。依据该向量的长度,生成上述四种变换矩阵,然后将它们合并生成为本文字典。显然,本文字典 Φ 的列数要远远大于行数,这就导致了该系统成为一个高度冗余的线性系统。由于深度图 x 的不同部分的不同形态适合字典 Φ 的不同原子来表达,该冗余系统更适合使用稀疏线性组合来表达目标值。

事实上,可以认为高分辨率深度图 x 由两部分组成:第一部分是观测值 $x_{obs} \in \mathbb{R}_m = \{x_i\}, i \in I_{obs}$;第二部分是缺失值 $x_{miss} = \{x_i\}, i \in I_{miss}$ 。这里的 I_{obs} 和 I_{miss} 分别是观测值和缺失值在深度图 x 中的索引。观测值 x_{obs} 是已知的,而缺失值 x_{miss} 是未知的,它是本文需要求解的对象。 x_{obs} 与 x 关系为:

$$x_{obs} = Sx \quad (5.2)$$

其中 $S \in \mathbb{R}^{m \times n}$ 是采样矩阵。因此, x_{obs} 可以表示为:

$$x_{obs} = Sx = S\Phi\alpha + \epsilon_{obs} = A\alpha + \epsilon_{obs} \quad (5.3)$$

系数向量 α 的稀疏性是非常重要的。如果不需要字典中的某个原子就能很好地表达当前深度图,那么 α 中对应的元素就是应该置为零。相反,如果有需要, α 中对应的元素可以是较大的值。Candès^[21]指出,只有 A 的相关性(coherence)低于某个值时,等式 5.3 才能唯一确定稀疏向量 α 。为了尽可能降低采样矩阵 S 和字典 Φ 之间的相关性,本文将表达 x 的字典修正为:

$$x = G\Phi\alpha + \epsilon = \Psi\alpha + \epsilon \quad (5.4)$$

其中, G 为高斯矩阵^[22]。在讨论等式 5.3 时,要求 α 是稀疏的。因此,从等式 5.4 中可知,高分辨率的深度图 x 可以由字典 Ψ 的各原子稀疏线性表示。

在缺失值 x_{miss} 存在的情况下, x_{obs} 中提供的信息不足直接求得系数向量 α 。本文采用最大期望算法(Expectation Maximization, EM)对该问题进行求解。EM 算法在迭代的过程中,逐步重建出缺失值和稀疏向量 α ,实现对最优化问题的求解。算法 5.1 给出了求解过程。

算法 5.1: 求解稀疏系数向量 α^{opt} Require: 观测值 $x_{obs} = Sx$, 系数向量的初始值 α_{old}

```

1: while(true)
2: 填充深度图  $x$  中的缺失值:  $x_{miss} \leftarrow (\Psi \alpha_{old})_{miss}$ 
3: 求解  $\alpha^{new}$ :  $\alpha^{new} = \arg \min_{\alpha} (\|x - \Psi \alpha\|_2^2 + \lambda \|\alpha\|_1)$ 
4: if(  $\|\alpha^{new} - \alpha_{old}\|_2 < \xi$ )
5:     break;
6:  $\alpha_{old} \leftarrow \alpha^{new}$ 
7: end
8: return  $\alpha^{new}$ 

```

5.2 实验结果与分析

本文实现了上述算法,在 Middlebury 标准测试集^[23-25]上进行了上采样测试。由于真实值 (Ground Truth) 的存在,本文对实验结果进行了量化研究,并与其它几个典型算法的结果进行了比较分析。然后,在用 Kinect 真实拍摄的深度图上,与一些非常典型的算法进行了视觉比较。

本文将从 Middlebury 测试集中随机挑选了 8 个场景进行对比测试。为了准确评价各个算法,本文采用了三种不同的方法来度量深度图的重建质量:1) PSNR (Peak Signal-to-Noise Ratio): 度量整体质量,它被认为是最接近人类视觉感知的图像质量度量^[26]; 2) DISC: 度量不连续区域的质量^[27]; 3) SRMS: 度量光滑区域的质量^[28]。从表 5.1, 表 5.2 和表 5.3 中可以看到,本文算法得到的结果要好于其他算法。

表 5.1 典型上采样算法在 Middlebury 数据集测试所得的 PSNR (Peak Signal to Noise Ratio, 单位: dB), 红色粗体数据是最优值。

场景	JBU ^[6]	NAFDU ^[7]	HQDMU ^[11]	ATGV ^[12]	JABDU ^[8]	本文算法
Tsukuba	24.7070	24.7230	27.1566	25.7923	19.9521	31.0249
Venus	35.4092	35.6574	34.0674	34.3038	37.5094	40.2136
Cones	34.8973	35.5132	38.0134	37.4503	30.3934	39.1856
Art	38.6320	38.9196	39.5957	38.3369	33.4176	42.8061
Books	32.9319	33.0280	33.4420	32.1950	25.0264	35.1577
Laundry	39.5784	39.9109	34.7523	36.0453	31.6916	42.8052
Aloe	25.5634	25.5480	27.2368	25.9905	22.9660	32.9223
Bowling1	26.8014	26.9376	27.5619	27.0297	21.9769	28.7485

表 5.2 典型上采样算法在 Middlebury 数据集测试所得的 DISC(阈值为 3),红色粗体数据是最优值。该数据表明本文算法在大多数情况下表现得最好。

场景	JBU ^[6]	NAFDU ^[7]	HQDMU ^[11]	ATGV ^[12]	JABDU ^[8]	本文算法
Tsukuba	0.298000	0.281174	0.545709	0.524584	0.413348	0.563283
Venus	0.218409	0.197820	0.249092	0.345983	0.375454	0.196609
Cones	0.249012	0.224992	0.401338	0.401338	0.399209	0.134387
Art	0.315813	0.300795	0.415194	0.393110	0.470848	0.195230
Books	0.225321	0.207706	0.251009	0.319633	0.354128	0.114495
Laundry	0.216505	0.194175	0.303883	0.444660	0.466990	0.143689
Aloe	0.309127	0.273798	0.420805	0.322866	0.340530	0.272620
Bowling1	0.317683	0.282927	0.584756	0.502439	0.400610	0.157927

表 5.3 典型上采样算法在 Middlebury 数据集测试所得的 SRMS,红色数据是最优值。该数据表明本文算法在大多数情况下表现得最好。

场景	JBU ^[6]	NAFDU ^[7]	HQDMU ^[11]	ATGV ^[12]	JABDU ^[8]	本文算法
Tsukuba	8.09088	8.00476	5.38272	7.24882	21.8286	2.09826
Venus	3.25756	3.14232	4.57377	4.33281	2.22988	1.86933
Cones	3.05849	2.93161	1.76789	1.68073	7.03863	1.72868
Art	2.39792	2.28305	2.18021	2.59023	4.79732	1.69144
Books	4.45554	4.42903	4.62262	5.22905	13.8256	4.34348
Laundry	2.33214	2.22414	4.35383	3.74181	6.37146	1.75583
Aloe	10.4427	10.6131	8.36018	10.1975	16.0206	2.63442
Bowling1	9.89235	9.86567	8.65488	9.36677	18.4914	9.23894

本文算法也应用于重建真实场景的高分辨率深度图。本文算法首先将低分辨率深度图投影到高分辨率的彩图上(图 5.1)。这些深度图含有大量的噪声,而且投影后的深度图并不是均匀的,有些地方含有空洞。重建这样的深度图是非常具有挑战性的。本文也分别应用了 HQDMU^[11]算法和 Ferstl 等人提出的 ATGV 算法^[12],图 5.1 给出了与本文算法结果的对比。从这些结果可以看出,本文算法给出的结果是非常好的。

6 结论

三维重建在科研和工业上都有重要用途,本文针对当前获取深度信息方法的种种缺陷,提出了一种新颖的低分辨率深度图上采样算法。本文算法克服了传统上采样算法过分依赖彩色纹理的客观缺陷,运用压缩理论等知识,在彩图分割的基础上分区域对深度图进行重建。整个重建过程摆脱了彩色纹理的干扰,避免了纹理复制这个问题。实验表明,本文算法可以很好地处理深度图中的噪声和空洞等问题,无论是视觉效果还是量化指标,本文算法都要优于其他算法。另外,本文还对当前彩色相机和深度相机的立体标定技术做了研究,并提出了深度图中矩形平板的自动提取算法,使深度相机的标定

更加自动化。本文也对图像分割技术做了相关研究,提出了一种可以联合低分辨率深度图和高分辨率彩图的区域合并算法,可以将场景分割得更加准确。



图 5.1 三个不同算法结果的对比。第一行的图片分别为两个不同场景的彩图和低分辨率深度图。不同场景的实验结果放置在不同的黑色包围盒中。从左至右,依次为 HQDMU^[11]、ATGV^[12] 和本文算法生成的深度图和含有彩图纹理的三维点云。在第一个场景中低分辨率深度图没有完全覆盖场景中的矩形平板,且矩形平板与墙面存在深度图差(绿色框)。在第二个场景中,打印机的纹理与墙面一样,但两者的深度是不同的(黄色框)。可以看到,本文算法重建的深度图更为准确。

参 考 文 献

- [1] Heiko Hirschmuller, Daniel Scharstein. Evaluation of Cost Functions for Stereo Matching[C]. //IEEE Confer-

-
- ence on Computer Vision and Pattern Recognition. IEEE, 2007:1–8
- [2] ean-Luc Starck, Michael Elad, David Donoho. Redundant multiscale transforms and their application for morphological component separation[J]. *Advances in Imaging and Electron Physics*, 2004. 132:287–348
- [3] Marco F Duarte, Yonina C Eldar. Structured compressed sensing: From theory to applications[J]. *Signal Processing*, IEEE Transactions on, 2011. 59(9):4053–4085
- [4] Matthias W Seeger. Bayesian inference and optimal design for the sparse linear model[J]. *The Journal of Machine Learning Research*, 2008. 9:759–813
- [5] Antonio Torralba, William T Freeman. Properties and applications of shape recipes[C]//2013 IEEE Conference on Computer Vision and Pattern Recognition. volume2. IEEE Computer Society, 2003:383–383
- [6] Johannes Kopf, Michael F Cohen, Dani Lischinski, Matt Uyttendaele. Joint bilateral upsampling[C]//ACM Transactions on Graphics (TOG). volume 26. ACM, 2007:96
- [7] Derek Chan, Hylke Buisman, Christian Theobalt, Sebastian Thrun, et al. A noise-aware filter for real-time depth upsampling[C]//Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008. 2008
- [8] Joohyeok Kim, Gwanggil Jeon, Jechang Jeong. Joint-adaptive bilateral depth map upsampling[J]. *Signal Processing: Image Communication*, 2014. 29(4):506–513
- [9] Andrew Adams, Jongmin Baek, Myers Abraham Davis. Fast high-dimensional filtering using the permutohedral lattice[C]//Computer Graphics Forum. volume 29. Wiley Online Library, 2010:753–762
- [10] James Diebel, Sebastian Thrun. An application of markov random fields to range sensing[C]//Advances in neural information processing systems. 2005:291–298
- [11] Jaesik Park, Hyeongwoo Kim, Yu-Wing Tai, Michael S Brown, Inso Kweon. High quality depth map upsampling for 3d-tof cameras[C]//Computer Vision (ICCV), 2011 IEEE International Conference on. IEEE, 2011: 1623–1630
- [12] David Ferstl, Christian Reinbacher, Rene Ranftl, Matthias Rüther, Horst Bischof. Image guided depth upsampling using anisotropic total generalized variation[C]//Computer Vision (ICCV), 2013 IEEE International Conference on. IEEE, 2013:993–1000
- [13] C Herrera, Juho Kannala, Janne Heikkilä, et al. Joint depth and color camera calibration with distortion correction[J]. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 2012. 34(10):2058–2064
- [14] Carolina Raposo, Joao Pedro Barreto, Urbano Nunes. Fast and accurate calibration of a kinect sensor[C]//2013 International Conference on 3DTV—Conference. IEEE, 2013:342–349
- [15] Shachar Fleishman, Daniel Cohen-Or, Cláudio T. Silva. Robust moving least-squares fitting with sharp features[J]. *Transactions on Graphics*, 2005. 24(3):544–552
- [16] Pablo Arbelaez. Boundary extraction in natural images using ultrametric contour maps[C]//Computer Vision and Pattern Recognition Workshop, 2006. CVPRW’06. Conference on. IEEE, 2006:182–182
- [17] David R. Martin, Charless C. Fowlkes, Jitendra Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues[J]. *Pattern Analysis and Machine Intelligence*, IEEE Transactions on, 2004. 26(5):530–549
- [18] Derek Hoiem, Andrew N Stein, Alexei A Efros, Martial Hebert. Recovering occlusion boundaries from a single image[C]//Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, 2007:1–8
- [19] Stéphane Mallat. Geometrical grouplets[J]. *Applied and Computational Harmonic Analysis*, 2009. 26(2):161–180
- [20] Emmanuel Candes, Laurent Demanet, David Donoho, Lexing Ying. Fast discrete curvelet transforms[J]. *Multi-scale Modeling & Simulation*, 2006. 5(3):861–899

-
- [21] Emmanuel J Candès, Michael B Wakin. An introduction to compressive sampling[J]. Signal Processing Magazine, IEEE, 2008, 25(2):21–30
 - [22] Jianchao Yang, John Wright, Thomas S Huang, Yi Ma. Image super-resolution via sparse representation[J]. Image Processing, IEEE Transactions on, 2010, 19(11):2861–2873
 - [23] Scharstein, Daniel and Richard Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms[J]. International Journal of Computer Vision, 2002, 47(1–3):7–42
 - [24] Scharstein, Daniel and Richard Szeliski. High-accuracy stereo depth maps using structured light[J]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003, 1:I–195–I–202
 - [25] Scharstein Daniel and Chris Pal. Learning Conditional Random Fields for Stereo[C]. //IEEE Conference on Computer Vision and Pattern Recognition, IEEE, 2007:1 – 8
 - [26] Huynh-Thu Quan and Mohammed Ghanbari. Scope of validity of PSNR in image/video quality assessment[J]. Electronics Letters, 2008, 44(13):800–801
 - [27] Daniel Scharstein, Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms[J]. International journal of computer vision, 2002, 47(1–3):7–42
 - [28] Ming-Yu Liu, Oncel Tuzel, Yuichi Taguchi. Joint geodesic upsampling of depth images[C]//Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013:169–176

Research of the Algorithm for Three-dimensional Reconstruction Based on Stereo Vision

Kong Dexing

(School of Computer Science and Technology, North China Univ. of Tech., Beijing, 100144)

Abstract Three-dimensional reconstruction is very important in computer vision and computer graphics. It indicates the development of computer technology of a country and has a lot of applications in academia and industry.

Capturing accurate depth information of a scene is the most critical content in three-dimensional reconstruction. In this paper, we propose a novel method that constructs a high-resolution depth map with high quality from a low-resolution depth image that is noisy and contains holes. We believe that sparse linear combination of atoms from an over-complete dictionary generates the high-resolution depth map, and the low-resolution depth map are the samples from the high-resolution depth map.

The proposed method is divided into three steps. In the first step, we stereo calibrate a color camera and a depth camera so that the low-resolution depth map can be projected onto the high-resolution color image. In order to calibrate depth camera automatically, we propose a rectangular plane extraction algorithm. In the second step, we segment the scene to find the regions with continuous changing depth by combining the color texture information with low-resolution depth map. In the last step, for each region, we find an optimal high-dimensional sparse vector representing the region best in Bayesian optimization framework, and then reconstruct the whole high-resolution depth map.

We implement the proposed algorithm, and apply it on Middlebury dataset and real scenes.

Comprehensive quantitative comparisons show that our method outperforms existing approaches when applied on Middlebury dataset, and qualitative comparison on real scenes indicates that our algorithm performs best

Key words three-dimensional reconstruction; camera calibration; image segmentation; sparse representation

基于连通域的扭曲文本图像校正方法研究与应用

研究生 郭正东

导师 曾凡锋

(北方工业大学计算机学院 北京 100144)

摘 要 OCR(Optical Character Recognition, 光学字符识别)

本文将致力于解决图像失真中的扭曲失真问题,通过对扭曲图像和中文文字结构特征的分析,总结国内外近阶段在扭曲校正问题上的优秀研究成果,对比分析各种方法的优缺点,适当参考并创新研究,针对中文文本图像提出了基于连通域的扭曲图像校正算法。

首先对图像进行预处理,包括灰度化、二值化、去噪和裁边等,然后进行连通域搜索,再通过字行交叉定位法提取文字和文本行,利用这些提取信息进行扭曲预判和量化,最后对扭曲图像进行重构,得到校正图像。

实验结果表明本课题解决方案能够有效得解决扭曲文本图像带来的识别问题,而且在方法的实现效率上有很好的控制,对其在实际项目中进行应用有着广阔的前景和不错的价值。

关键词 中文文本图像; 扭曲图像; 连通域; 文字切分; 字行交叉定位法

作者简介 郭正东,男,1990 年出生,硕士研究生,主要研究方向为图像处理。

E-mail: guo. zhengdong@163. com

1 绪论

近几十年来,图像的扭曲校正技术获得迅速发展,从校正对象可以分为两个主要方向:(1)基于 3D 的模型重建技术校正(2)基于 2D 图像校正,目前二维图像校正技术又可以分为:a. 基于连通域处理,一直以来都是该研究领域主流,不过由于处理细致,也导致了这类方法的速度较慢,亟待提高。b. 基于文本线处理,这类方法增大了校正粒度,减少处理对象,提高了处理速度,但也是以行为最小校正单位进行处理,所以对校正的精准度有所损减。c. 基于区域信息校正,这类方法能够达到快速而且较好的校正效果,只是通常对原始图像拍摄方式有特殊要求,适用范围受到限制。

经过上述分析,目前基于 2D 的扭曲校正方法各有优缺点,其中许多经典的校正方法是针对英文文本图像设计的,受文字结构差异影响,应用到中文图像时获得的效果并不理想,难以直接套用,因而需要研究其他适应性更强的方法来完成校正。

2 基于连通域的扭曲校正算法实现

本方法主要思路是先对图像进行预处理,包括灰度化、二值化、去噪和裁边等,然后进行连通域搜索,在通过字行交叉定位法提取文字和文本行,利用这些提取信息进行扭曲预判和量化,最后对扭曲图像进行重构,得到校正图像。

2.1 文字分割算法

本文参考传统方法,分析了扭曲文档中文字的特征,结合中文印刷品排版的特点提出了一种优化的快速文字提取的方法,即字行交叉定位法,先为标记的连通域添加包围边框,然后按给出条件分成多步合并连通域,逐步将不同结构类型的文字的文字提取出来。

2.1.1 添加基本连通域包围边框

为连通域添加包围边框是文字合并的前提,也是对文字的一次粗提取,汉字结构复杂多样,多数都排布紧凑,无论是上下结构或者左右结构都频繁出现了笔画粘连的情况,这为为文字提取带来了极大的便利,因为,仅仅对初始连通域添加边框即可将多数文字提取出来,具体算法如下。

创建连通域边框数组 R ,上下左右值初始化为-1,遍历连通域标号数组 $merge$,比较同一连通域各点坐标(记为 $p.x$ 和 $p.y$)和 $R[merge[x, y]]$ 的上、下、左、右四个属性值(记为 $R[p].l$,同理于 t, r, b),按下列公式更新 R 数组取值,即:

$$R[p].l = \begin{cases} p.x & \text{当 } p.x < R[p].l; \\ R[p].l & \text{否则不变。} \end{cases} \quad \text{公式(2-1)}$$

$$R[p].t = \begin{cases} (height - p.y) & \text{当 } (height - p.y) < R[p].t; \\ R[p].t & \text{否则不变} \end{cases} \quad \text{公式(2-2)}$$

$$R[p].r = \begin{cases} p.x & \text{当 } p.x > R[p].r; \\ R[p].r & \text{否则不变。} \end{cases} \quad \text{公式(2-3)}$$

要注意的是,计算边框时是根据视图的窗口坐标,原点在左上角,而图像的数据起点是从图片的左下角,因此需要如公式(4-4)进行变换。另外,由遍历顺序可知,边框下界 $R[p].b$ 不可能出现更大值,初始赋值为 $(height - p.y)$ 即可,其中 $height$ 代表图片的高度。

2.1.2 文字的重叠合并和纵向相邻合并

在上步之后,文章中结构紧凑的多结构文字和单一结构文字都已提取出来,下一步开始合并离散结构的文字。这类文字是由多个连通域组成,为了将同一文字的各结构归并到同一连通域当中,这里用到了重叠合并和近邻合并两种方法连续处理,算法如下:

第一步确定图像中字体的标准宽高,先统计连通域边框的宽高,然后寻找出现概率最大的高度 H_m ,因为多数文字已经被正确分割出来,且一般的印刷设计版面中大部分的文字高度相差无几,所以可将 H_m 作为标准高度 H_s ,考虑到字体变化和英文字母等干扰因素,标准高度的估计从最大高度的 $1/3$ 处开始寻找更加准确。标准宽度的确定不能照用确定 H_s 的方法,因为文档中英文单词、标点符号、数字等非汉字元素的宽度变化较大,另外,当前处理的是扭曲图像,扭曲部分文字的宽度也会缩小,这些原因导致了宽度分布复杂。观察发现,利用印刷版面的特征发现,通常具有固定的宽高比例标准,通过标准高度来计算标准宽度 W_s 。

第二步循环合并,文字中的包围结构和半包围结构由于其连通域包围边框产生重叠,因而可以通过重叠合并进行提取,上下结构的文字,其连通域不存在重叠,合并方法采用的是相邻合并,但是左右结构的文字不能直接进行相邻合并,因为在扭曲图像中,文字会受到横向的压缩变窄,直接进行相邻合并可能将多个文字并入同一连通域中,另外,文中的数字、字母、标点符号等也会影响到横向合并,所以,左右结构的文字需要稍后处理。

按序遍历连通域 C_i ,为当前连通域周围寻找其他连通域进行合并时,建立一个待合并区域搜索窗口 Win 能够大大提高效率,免去了大范围的搜索,窗口 Win 创建条件如下:

$$Win.left = R[C_i].left \quad \text{公式(2-4)}$$

$$Win.right = Win.left + W_s \quad \text{公式(2-5)}$$

$$Win.top = R[C_i].center - H_s \quad \text{公式(2-6)}$$

$$Win.bottom = R[C_i].center + H_s \quad \text{公式(2-7)}$$

当搜到其他连通域 C_j 时,按下列条件判断能够进行重叠合并:

$$H_{com} = \max(|C_{i,b} - C_{j,t}|, |C_{i,t} - C_{j,b}|) \quad \text{公式(2-8)}$$

$$W_{com} = \max(|C_{i,r} - C_{j,l}|, |C_{i,l} - C_{j,r}|) \quad \text{公式(2-9)}$$

$$(W_i + W_j) / W_{com} \geq 1 \text{ 且 } (H_i + H_j) / H_{com} \geq 1 \quad \text{公式(2-10)}$$

若符合条件表明发生重叠则进行合并,否则,将其编号 C_j 记录到相邻连通域集合 S 中,每次仅处理宽高相加大于 2 的连通域,此做法有一定的逻辑去噪作用。

当遍历窗口 W 完成后,集合 S 若不为空,逐一遍历其中的剩余连通域,尝试进行相邻合并,判断条件为:

$$(W_i + W_j) / W_{com} \geq 1 \quad \text{公式(2-11)}$$

$$\rho \leq (H_i + H_j) / H_{com} < 1 \quad \text{公式(2-12)}$$

$$W_{com} \leq W_s \text{ and } H_{com} \leq H_s \quad \text{公式(2-13)}$$

ρ 是限定合并条件的重要参数,若 ρ 过大会导致合并不完整使得同一个文字的偏旁部首不能并入同一个连通域, ρ 过小则又可能导致合并过度,使得多个文字或标点符号等非文字元素被并入一个连通域中。经过多次实验测试,当 ρ 取值在(0.29,0.32)之间时有很好的文字切分。

按照以上条件进行两次搜索判断并不能完全达到我们的要求,因为会出现第一次判断合并时没有满足条件,第二次却满足了条件,因为每次合并后连通域区间都会增大,之前不重叠或者不相邻的连通域可能发生了重叠或者足够小的邻近。因此,需要不断的检查待合并连通域集合中的连通域,直到没有连通域符合合并条件为止。至此就完成了将非左右结构的文字提取出来。

2.1.3 初步定位文本行

这里文本行定位的目的在于将同一行中的连通域左右相邻关系确定,通过比较相邻连通域的距离能够找到优先合并的连通域,因为在印刷排版中字内间距肯定小于字间间距,且如之前所说的,扭曲造成的文字变形和非文字元素的影响也决定了不能直接进行左右相邻合并。文本行的定位方法应用多次,具体方法将在下一节给出。

2.1.4 完成文字提取

文本行定位之后,以行为单位统计文字的左右邻居,记录相邻文字的边界间距,将间距小于给定阈值的连通域对结合间距,按照间距从小到大的次序存入集合 U 中,然后遍历集合元素,进行相邻合并,合并条件参考纵向相邻合并的条件,如下:

$$\rho \leq (W_i + W_j) / W_{com} < 1 \quad \text{公式(2-14)}$$

$$W_{com} \leq W_s \text{ and } H_{com} \leq H_s \quad \text{公式(2-15)}$$

经试验测试, ρ 的取值在(0.49,0.85)之间比较合适,这里 ρ 的取值主要影响的是操作速度,对合并结果影响不大。最终的合并效果如图 4-7 示例。

式系统开发,数据挖掘,电子
通信工程等领域取得重大成
新科研成果应用于“奥运会
70多项成果通过省部级鉴定
余篇,出版专著或译著 160
项,荣获 2004 年度国家科技
步二等奖 2 项。本项目依托

图 2-1 文字切分效果图

2.2 文本行定位算法

文本行搜索是本课题研究的另一创新点,由于拍摄图像扭曲文本线的走向随意,不宜采用水平投影分析来判断文本行,本文采用的是基于文献[8]提出的就近聚合文字的方法定位文本行,具体算法如下:

第一步计算各个文字的中心坐标(C_{m_x} , C_{m_y});

第二步就近连接左右邻居。依据中心坐标,计算 C_i 与左右两侧文字的中心间距,并分左右两边的各取一个间距最小的文字 C_{i_l} 和 C_{i_r} ,从印刷版面分析可以发现各字块(包含文字和如标点符号、字母和数字等独立字符)与左右字块的间距最小,因此, C_{i_l} 和 C_{i_r} 为 C_i 的左右邻居字块。为提高效率,通过公式计算中还跳过明显不在一行的字块,对 C_i 和 C_j 而言, C_i 的上边界小于 C_j 的下边界,或 C_i 的下边界大于 C_j 的上边界时,认为 C_i 和 C_j 不在同一行。

第三步将文字聚合成行。先通过遍历各个连通域,寻找出各行的行首文字,没有左邻居的就是行首文字。再从行首文字开始,迭代查找右邻居直至找到没有有邻居的文字完成一行。

实验效果如图 4-8,通过连接各行文字的中心来验证文本行定位的正确性。

由于非财务指标与企业的战略规划的相关度较高,因此越来越受到管理人员的青睐与重视,其作为企业长期发展战略的一个终端,对企业的盛衰成败有很大的影响。因此,企业应当在了解本企业内外环境的基础上,结合非财务指标的优缺点,完善现有的业绩评价体系。当企业管理者在建立绩效评价体系时,有必要在总体上对绩效评价体系的构成做一些原则性的规定,以使其符合企业的长远发展。本文认为,企业对非财务指标体系的设置问题应至少符合以下原则:关联与独立并存的原则、相关性原则、可操作性强并兼顾成本效益原则。

同时,企业在使用非财务指标时应考虑以下几个方面的因素:第一,关于企业高管人员的变动;第二,企业自身各方面的特性;第三,非财务指标与财务信息的关系;第四,对非财务指标赋予的权重;最后是关于企业自身内部环境的影响。当然在设置过程中还应当考虑企业非财务数据的来源以及可量化性。对于大多数企业来说,对其业绩的最有效评价仅仅使用非财务指标来满足是不够的,应当设置成为一个完整的体系,既包含非财务指标,也包含财务指标,而且应当避免两类指标之间的冲突。

图 2-2 文本行定位示意图

2.3 扭曲预判与量化

扭曲预判和量化功能是为了本算法在实际应用中让产品使用更加智能化而设计的。扭曲预判功能能够判断扭曲是否存在,扭曲量化的设计是为了在扭曲预判发现扭曲后对扭曲预判的结果进行量化,通过实验评估扭曲图像有没有校正价值。

2.3.1 扭曲预判

本文扭曲预判的方法是结合了助视仪采集图像的特点,如图4—9。图像中扭曲图像的文本行具有一定的规律性,即图像出现扭曲的文本行会在扭曲端出现最高点或者最低点,在图像的上方扭曲会有最高点,图像下方的扭曲会有最低点,所以扭曲图像通常在扭曲端存在一条极值点连线,如左图,而在正常图像中,文本行平齐,排列规整,不会存在极值点,因此可以利用这一点判断扭曲,并能确定扭曲发生在左端或者右端。

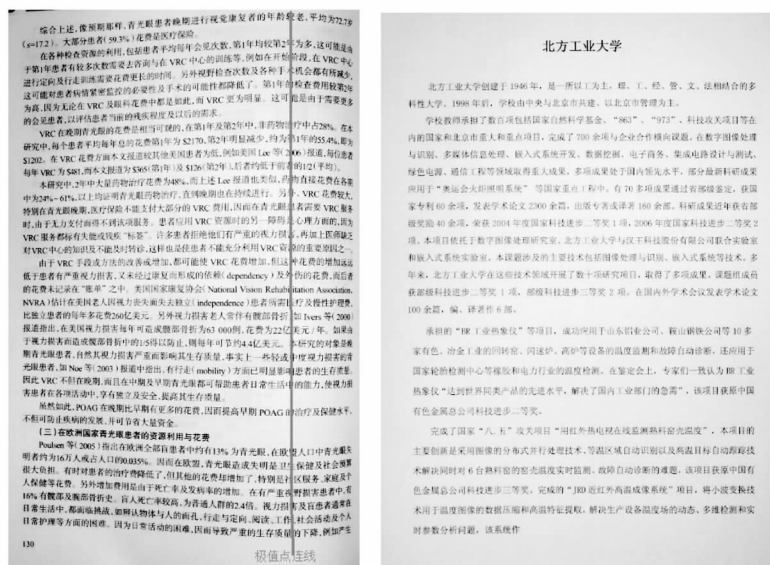


图 2—3 扭曲图像的极值点(左)与正常图像(右)

因此本文的扭曲预判方法是统计极值点的数量和相对位置,极值点超过行数 75%—80%时,必定存在扭曲,极值点所在位置就是扭曲发生的一侧。

2.3.2 扭曲量化

扭曲量化并没有具有完整结论,需要先对文本线进行曲线拟合,下列指标是扭曲量化的参考参数:

- 极值点的曲率最大值,或者纵向位置行差 K ;
- 文本行纵向占用高度差 H 与文字标准高度 $height$ 的比值 r ;

本课题的量化公式是:

$$\theta = r \cdot \sqrt{\frac{\sum K^2}{n}} \quad \text{公式(2—16)}$$

经多次探索,该公式能够基本描述出扭曲程度变化的趋势,与原始图像的识别率保持同步。

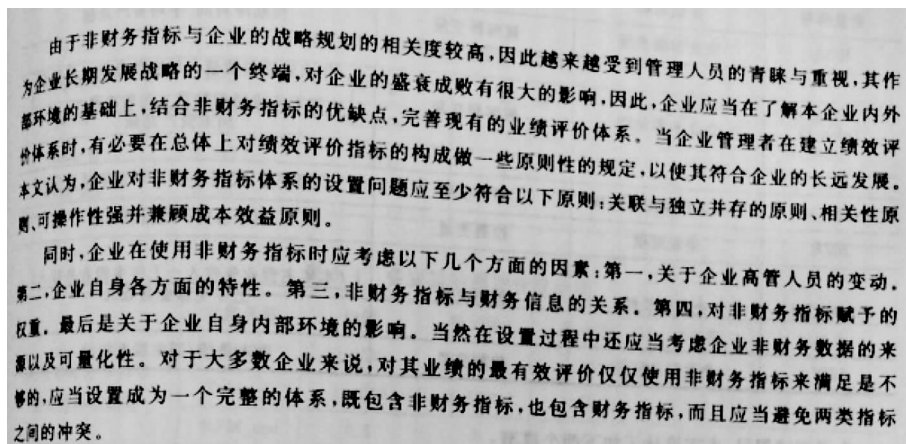
2.4 图像校正方法

前面已经确定了各行之间及行内文字之间的逻辑位置,并获得了各文字的实际坐标区间,由于各个汉字的所占面积相接近,那么按行将所有文字移到合理位置即可达到扭曲校正的目的。具体操作

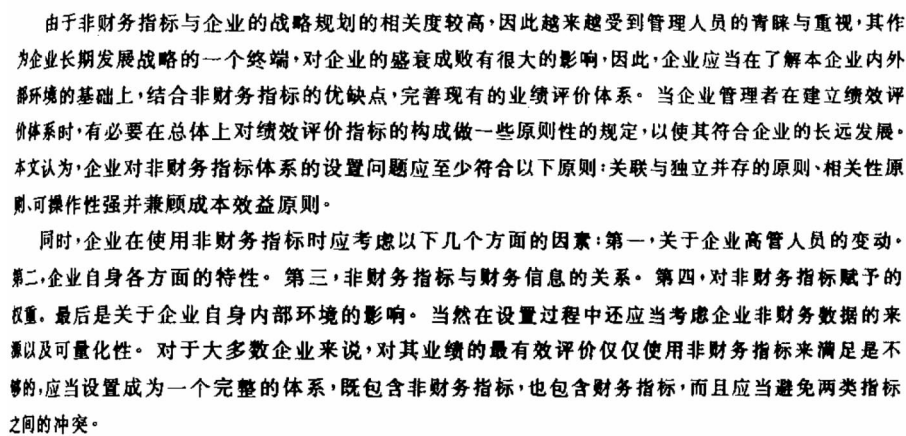
是:对图像由上到下逐行遍历,以第 i 行 $L_i (i > 0)$ 为例,先以 L_i 的最左文字作为标记文字 C_m ,然后遍历其他文字 $C_n (n > 0)$,按公式(4-17)计算校正位移 d ,

$$d = C_{m-y} - C_{n-y} \quad \text{公式(2-17)}$$

最后将 C_n 按校正位移 d 对文字垂直移动(d 的正负值对应于上移和下移),并绘在新图像中,因为如果在原图上操作,可能会使相邻两行出现重叠,因为 C_n 可能占到了 C_m 的校正位置。使得各行文字校正同一水平线上就完成了扭曲文档图像的文本行扭曲校正。效果图如图 2-4。



(a) 原图



(b) 校正后图像

图 2-4 校正效果图

3 校正方法评测与实验结果分析

为了检验效果,同时选取了两种近年来 2D 扭曲校正方向有一定成果的方法进行对比。实验结果如表 3-1。

表 3-1 各算法校正耗时对比

	文献[1]	文献[2]	文献[3]	本文算法
平均耗时(ms)	>3000	>3500	>1000	813

可以得到,从耗时角度上看,本文方法缩短了校正时间,较好地控制了基于连通域校正方法的弊病。然而衡量本文方法价值的一个更为重要的标准是校正结果的识别率,下面将对三种方法校正后图像识别数据的统计结果。

本节统计了三种方法校正图像后的识别效果,通过将图像按原始样张识别率作为扭曲程度参考进行分类,对各种程度的扭曲图像分别进行统计,获得的结果如表格 3-2。

表 3-2 各算法校正效果对比

	原始样张	文献[1]	文献[2]	文献[3]	本文算法
不同扭曲程度样张 的校正后识别率	≤50%	50%以上	80%以上	90%	93%以上
	(50%,80%)	94.2%	90.6%	93.0%	94.5%
	≥80%	95.3%	93.6%	94.5%	95.9%

从校正效果上看,对于扭曲程度较高即识别率不足 50%的图像,文献[1]方法提取文本行时局限于假设了页边区域文本行水平进行处理,因而提取文本行不准确导致识别率偏低,但其他情况都有较好的校正效果。文献[2]和[3]的处理能力较为稳定,具有较好的校正作用,而本文的方法在识别率上能够在不同扭曲程度的校正实验中均有明显效果,相比而言,较有优势。

4 总结

通过比较三种方法的校正耗时和不同扭曲程度样张的校正后识别率作为评判依据,通过以上结果分析表明,本文方法对文字处理细致,速度快且仍能保证较高的识别率,适用范围较大。本方法的算法设计已进行多次优化,尤其是文字切分部分,之前尝试不同的算法由于遍历次数较多影响速度,后经改进,设置窗口并用遍历连通域编号来代替像素遍历,耗时得到了改善。

参 考 文 献

[1] Liu Hong, Ye Lu. A method to restore Chinese warped document images based on binding characters and building curved lines [C]. International Conference on
[2] 宋丽丽,吴亚东,孙波. 改进的文档图像扭曲校正方法,计算机工程[J]. 2011,37 (1):204-206
[3] 曾凡锋,王晓,吴飞飞. 基于文本行重构的扭曲文档快速校正方法,计算机工程与设计[J]. 2014,35(2):573-577

The Research and Application of Correcting Method Based on Connected Components for Warped Chinese Document Images

Guo Zhengdong

(School of Computer Science and Technology, North China Univ. of Tech. , Beijing, 100144)

Abstract This paper will commit to solve the problem of warped distortion in image. By the analysis of warped image and the characters feature, the summarize of many excellent research achievements from domestic and abroad in recent years, and compared the advantages and disadvantages of those methods, a fast correcting method has been proposed based on connected components for warped Chinese document images.

Firstly, the image should be preprocessed, such as graying, binarization, denoising, image cutting and connected components searching. Then searching connected components, and extracting characters and locating lines, that is Characters and Text lines Locate Alternately(CTLA). After that judging and quantitating the exist of warping and by used all located information, Finally the warped image was correcting by reconstruct it.

Lastly, all introduction and explanation indicate this solution can solve warped distortion in Chinese document image with nice time-consuming control, that shows its valuable practical applications.

Key words chinese document image; warped image; connected components; character segmentation; characters and text lines locate alternately

高速低功耗无线传感器网络 数据传输技术研究与设计

研究生 高洪磊

导师 马 礼

(北方工业大学电子信息工程学院 北京 100144)

摘 要 无线传感器网络一般用于传输温湿度等信息量小的数据,对于传输图像、音视频等多媒体数据时,传输速率就受到关注,也进一步突出了其节点能耗和信息传输时效性问题。传感器网络协议栈中 MAC 层直接控制着无线射频模块,其协议设计直接影响着节点的能耗以及数据传输时效性,是传感器网络研究的重要内容。

本文针对传统无线传感器网络在传输大数据量的多媒体信息时面临的挑战,在分析 MAC 协议的基础上,从异步 MAC 协议角度开展研究。

现有的基于接收端发起的异步 MAC 层协议,在预测接收节点唤醒时间时,多采用维持邻居节点的伪随机唤醒时间表的方法。在传输多媒体信息的应用中,网络负载波动较大,负载动态变化,伪随机唤醒时间表方法不能动态的改变节点唤醒间隔,从而带来较高的数据传输冲突率及较大的数据传输延时。为了解决此问题,提出一种高能效、低延迟协议 HELD-MAC(High energy efficiency and Low delay MAC),保证节点之间具有不同的唤醒时间,同时实现准确预测接收节点唤醒时间,而且接收节点能够根据网络负载情况动态改变唤醒间隔。同时为了减轻能量黑洞问题,协议根据节点的剩余能量,自适应地改变节点的最小唤醒间隔。通过仿真对 HELD-MAC 协议性能进行评估,与 RI-MAC、PW-MAC 等协议进行对比,在数据传输延时、能量消耗、网络吞吐量、传输碰撞方面具有明显优势。HL-MAC 协议的平均占空比、平均数据传输延时比 RI-MAC 少 81.2%、19.5%,比 PW-MAC 少 11.2%、10.1%。网络生存周期比 RI-MAC 延长约 9.5%,比 PW-MAC 高 4.4%。

HELD-MAC 相比其协议更加适合传输大数量的以及对传输时效性要求较高的无线传感器网络应用。

关 键 词 无线传感器网络;多媒体信息;MAC;异步

作者简介 高洪磊,男,1990 年出生,硕士研究生,主要研究方向为无线传感器网络。

E-mail: crazy_leilei@126.com

1 绪论

WSN 融合了多项技术,包括无线传输技术、分布式计算技术、嵌入式技术等,它是由大量的经过自组网形成的多跳路由传输的节点构成,用于对监测区域内的感知对象进行感知,然后对其进行采集,经过多跳传输的形式发送到观察者。随着传感器网络技术的发展,WSN 已经得到广泛的应用。

WSN 体系结构通常包括传感器节点、网关节点和用户管理节点,其中网关节点又被称为 Sink 节点和汇聚节点。传感器节点通常被散布到监测区域内,以自主组网方式形成网络,节点之间形成链路。节点感知数据并对其进行处理后,通过多跳转发到汇聚节点,到达汇聚节点之后,根据汇聚节点具有的通信功能,通过互联网或者其它方式中继到用户节点,用户节点对汇聚的数据进行集中处理、存储和展示。

WSN 体系结构通常包括传感器部分、处理器部分、无线通信部分和电量供应部分。

传感器节点和汇聚节点协议栈使得节点收发功率和路由组网得到有效融合,兼备网络协议的数据融合,利用无线传输方式进行高效的通信,促进了传感器节点之间高效协同合作。

在 WSN 特点和应用限制中,传感器节点的能量约束是最重要的限制,因为节点的功耗的大小直接决定着传感器网络的生存周期,同时也直接影响着网络的拓扑结构,尤其是应用在无法长时间停留,不能更换传感器电池的应用环境中。因此,设计合理的 WSN 协议栈尽最大程度降低能量消耗,延长网络的寿命有着重要的研究意义。

本文将在研究异步 MAC 层协议的基础上,分析已有的异步 MAC 层协议的优劣势,基于经典的基于接收端发起的异步 PW-MAC 层协议的精华,在需求传输图像、音视频等特殊应用背景下,对 PW-MAC 协议进行优化改进,设计了一种新的能量高效、数据传输时延低的基于接收端发起的异步 MAC 层协议。

2 正文

2.1 高速传感器网络 MAC 层协议分析

HELD-MAC 协议主要包括以下几个环节:具有传输音频、视频、图像等多媒体信息的传感器网络,被称之为无线多媒体传感器网络(WMSN),传感器节点通常装备摄像头、麦克风等。WMSN 节点在节点部署初始化阶段,以自主组网的方式形成网络拓扑结构,感知和采集监测区域内多媒体数据并将其传输到下一节点。WMSN 作为 WSN 的一种,和传统传感器网络拥有相同之处,但是又同时具有其各自的特点。表 2-1 描述了 WSN 和 WMSN 特点比较。相似之处主要表现在大规模、自主组网、动态性、节点能量和通信能力受限等。其显著特性主要表现在以下几点:(1)可以处理大数据量的多媒体信息,节点计算和存储能力较强;(2)可以感知多种数据感知媒体丰富,多种异构数据共存;(3)处理任务复杂;(4)可以对环境进行全面有效感知^[2]

能耗问题是制约传感器网络发展的一个重要制约,过早的消耗节点能量会导致节点关闭,减少整个网络的生存周期。当某一监测区域内由于流量过大,消耗能量过多会导致大面积节点关闭,传感器网络覆盖区域受到影响,进而影响数据的准确性以及全面性。因此如何减少节点的能量消耗一直是 WSN 研究的重点内容。

基于传统传感器网络和 WMSN 的特性与区别,设计 MAC 层协议时面临新的挑战。设计 WMSN 的 MAC 层协议时,除了需要考虑传统 WSN MAC 层协议需求,例如能量消耗、串音、冲突处理、信道竞争等情况,同时由于多媒体信息量较传统数据数据量大得多,因此 MAC 层协议应该尽量有较高的吞吐量,同时必须考虑数据传输时效性,必须控制在一定的范围之内。基于以上两方面的需求,在设计 MAC 层协议时面临着新的挑战,主要表现在以下几点:

- (1)降低能耗
- (2)Qos 保障
- (3)区分服务
- (4)平衡各种性能

近十年来,大量的研究人员投入到无线传感器网络中 MAC 层协议的研究中,已经为传感器网络设计了大量的 MAC 协议。为了节省能量,尤其是空闲侦听带来的能量消耗,大多数协议采取的办法是将射频收发器尽可能的让其处于休眠模式,因为休眠模式消耗的能量比其他模式消耗的能量要少很

多,比如空闲侦听、数据包发送和接收等。在休眠模式,传感器节点不能够从信道中收发数据,这种方法很大程度上减少了由于空闲侦听带来的能量消耗。然而,为了节省能耗使节点长时间处于休眠状态必会带来数据传输延迟,进而影响数据传输速率,因此合理设计 MAC 层协议使节点低能耗和低传输延迟具有重要的研究价值。下面将从三种异步 MAC 协议出发,分析各种协议的节能策略和降低延迟策略。

(1) 预定时隙机制

(2) 同步 MAC 机制^[9-10]

(3) 异步 MAC 协议^[12-13]

本文将在研究异步 MAC 层协议的基础上,分析已有的异步 MAC 层协议的优劣势,基于经典的基于接收端发起的异步 PW-MAC 层协议的精华,在需求传输图像、音视频等特殊应用背景下,对 PW-MAC 协议进行优化改进,设计了一种新的能量高效、数据传输时延低的基于接收端发起的异步 MAC 层协议。节点之间具有相互独立唤醒/休眠时间,节点在能够准确预测邻居节点唤醒时间的同时,能够保证邻居节点之间具有不同的唤醒时间,从而保证了节点在不同时间唤醒传送数据,避免了节点在同一时间唤醒带来的数据传输冲突。考虑到需要传输图像、音视频等数据量较大的信息以及多出现数据量波动较大的情况,协议能够根据网络负载动态的、自适应地改变唤醒时间间隔,同时保证邻居节点仍能够准确预测邻居节点的唤醒时间。

2.2 低功耗高速传感器网络 MAC 层协议设计

HELD-MAC 协议主要包括以下几个环节:

(1) 新节点加入网络阶段。根据邻居节点唤醒时间选择偏移值,保证与邻居节点有不同的唤醒时间。在本章第四部分介绍的负载流量预估机制,节点的唤醒时间间隔随着网络负载情况自适应改变,因此本节设计的重点是如何确保参数的选取,保证节点之间始终具有不同的唤醒时间;

(2) 智能预测节点唤醒时间机制。为了减少发送节点因等待接收节点唤醒而保持唤醒状态带来的空闲侦听能量消耗,本部分提出了一种预测接收节点唤醒时间机制,此机制能够准确预测接收节点的唤醒时间,当发送节点有数据传输时,根据预测到的接收节点唤醒时间提前 T_{cd} 时刻唤醒,以等待接收唤醒信标帧;

(3) 节点自适应改变最小唤醒间隔。针对能量空洞问题,即转发节点由于传输的数据量较普通节点消耗的能量较多,可能过早的关闭的问题,提出了节点自适应改变最小唤醒间隔机制,根据节点剩余能量比率 V_{st} ,自适应地改变节点最小唤醒时间。此机制保证了节点不会因为由于能量过早消耗完毕,减少整个传感器网络的生存周期;

(4) 负载流量预估机制。针对特殊的传感器网络应用,即需要传输图像、音视频的应用,数据量较大,并且某一段时间数据量波动较大的情况,提出了负载流量预估机制。通过接收节点和发送节点通信机制,接收节点根据发送节点等待传输的数据包数量等级 ρ_{wait} ,自适应改变节点的唤醒时间间隔。当数据量增加时,节点减少唤醒时间间隔,当数据量减少时,节点增加唤醒时间间隔。此机制保证了大数据的及时传输,提高了数据的传输速率,减小了节点的能量消耗。

(5) HELD-MAC 协议传输冲突检测和数据重传机制。数据传输冲突与数据重传是传感器网络 MAC 层设计需要考虑的重要方面,如何减少数据传输冲突率和提高数据传输重传效率具有重要意义。

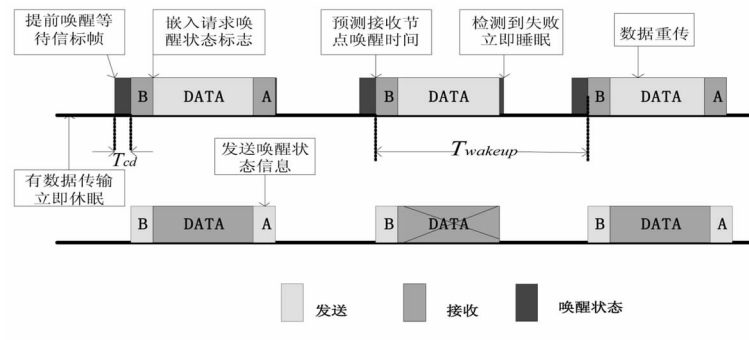


图 1-1 HELD-MAC 协议工作过程

2.3 协议仿真与实现

仿真实验在 WINDOWS 系统利用 VMware 软件虚拟出一套 Ubuntu 操作系统。然后在系统中运行 NS-2.34。

针对上面的参数配置完成之后,每次仿真之后都会得到 .tr 结尾的 trace 文件,此文件中包含了模拟仿真过程中协议栈各层发生的各种事件,包括节点的各种能量消耗,节点之间收发数据等。在本文中设计的仿真数据,都是来自仿真结束后生成的 trace 文件分析得到的,从这里看出来 trace 文件的重要性。

Set val (chan)	Channel/WirelessChannel	##信道类型
Set val(prop)	Propagation/TwoRayGround	##电波传输方式
Set val(netif)	Phy/WirelessPhy	##网络接口类型
Set val(ifq)	Queue/DropTail/PriQueue	##接口队列类型
Set val(mac)	Mac/HELD-MAC	##mac类型
Set val(ll)	LL	##链路层类型
Set val(ant)	Antenna/OmniAntenna	##天线模型
Set val(ifqlen)	50	##最大包长
Set val(nn)	100	##节点数
Set val(rp)	SCR	##路由协议
Set val(cp)	"cbr-5-test"	##数据流
Set val(sc)	"scen-6-s"	##场景文件
Set val(em)	EnergyModel	##能量模型
Set val(rx)	386mW	##接收数据能量消耗
Set val(tx)	368mW	##传输数据能量消耗
Set val(ie)	344W	##空闲状态能量消耗

图 2-1 配置脚本

仿真结束之后,生成了 trace 文件,此时就有了仿真过程的各种数据,接下来的重点就是针对生成的数据进行分析,由于生成的数据量很大,人工进行分析是很难实现的,针对此问题 NS2 自带了分析数据的 AWK 工具。AWK 脚本处理语言是一种强大的统计数据和模式匹配语言,具有使用方便、简单等强大优点。用户可以编写指定的匹配模式,然后去数据中匹配得到符合模式的数据。本文在进行数据分析时采用了 AWK 和 shell 语言相交互的方式。首先使用 awk 语言处理生成的数据,从生成的 trace 文件中摘取有用的数据,然后再用 shell 语言对提取出的有效数据进行进一步的处理和统计,最后计算出需要计算的性能指标。

(1) 实验一(随机网络实验)

在 900 m×900 m 的随机部署 49 个节点,在中心部署 1 个基站,节点通信半径为 200 m。假设节点间隔 0.8s 固定时间产生数据包,在某时间段内不出现数据包数量急剧增加的情况。仿真时间为

1h。

实验结束后,针对生成的 trace 文件,使用 awk 工具从大量数据中提取实验评估中需要的数据,包括每个节点收发包数量、各种状态持续的时间等信息,提取完成信息之后通过 shell 脚本完成数据统计。数据统计结果表 5-3 如下:

表 2-1 随机网络性能对比

协议	平均占空比/%	平均端对端传输延迟/s	平均碰撞次数/次
HELD-MAC	10.59	0.85	10.02
PW-MAC	15.89	1.05	12.5
RI-MAC	45.30	1.63	40.62
X-MAC	37.69	1.71	42.51

由表 5-1 可看出,在平均占空比方面,HELD-MAC 协议相比较与其他三种协议有明显优势,比 PW-MAC 协议减少 5.3%,比 RI-MAC 和 X-MAX 分别减少了 34.71%、27.1%。

HELD-MAC 协议的每个数据包传输的每跳平均能耗为 2.0,比 PW-MAC 协议减少了 0.5 mA,比 RI-MAC 和 X-MAC 协议减少了 1.2 mA 和 2.7 mA 因为在节点运作过程中,假设节点不会出现数据包急剧增加的情况,HELD-MAC 中节点的唤醒间隔大部分时间处在最大值,而 PW-MAC 中节点的唤醒间隔始终在 0.2 s—1.6 s 之间随机变换。同时 HELD-MAC 协议确保节点之间具有不同的唤醒时间,从而减少了数据传输冲突和数据重传。而 RI-MAC 和 X-MAC 不能准确预测节点的唤醒时间,需要每次醒来侦听信道,等待接收节点的唤醒信标帧,因此在平均占空比方面有较大劣势。

在平均端对端传输延时方面,HELD-MAC 比 PW-MAC 协议有略微提升,但是比 RI-MAC 和 X-MAC 有了较大提升。相比于 RI-MAC 和 X-MAC,HELD-MAC 和 PW-MAC 保证了节点之间不同的唤醒时间,减少了数据传输冲突和数据重传,从而减少了传输延时。PW-MAC 不能根据节点的负载情况,自适应地改变节点的唤醒间隔,因此靠近 SINK 节点的一些转发节点,由于需要转发的数据量较大,会带来传输冲突,从而带来数据传输延时。而 HELD-MAC 中节点可自适应改变节点的唤醒间隔,因此当检测到转发的数据量较多时,减小唤醒间隔,从而减小传输延时,同时 HELD-MAC 协议采用一次“握手”多次传输机制,进一步降低了数据传输延迟。

(2)实验二(网格网络实验)

在 4×4 网格网络中,SINK 节点处在网格中央,节点与邻居节点之间距离为 90m,节点通信半径为 100m。边缘节点固定时间间隔发送数据帧。通过改变此时间间隔来模拟网络中出现突发事件,导致数据量逐渐增加的情形。通过改变产生数据包时间间隔,观察 HELD-MAC、PW-MAC、RI-MAC 在应对网络负载增加时性能对比。实验结束后数据处理方法与实验一相同。

1) 能耗性能对比

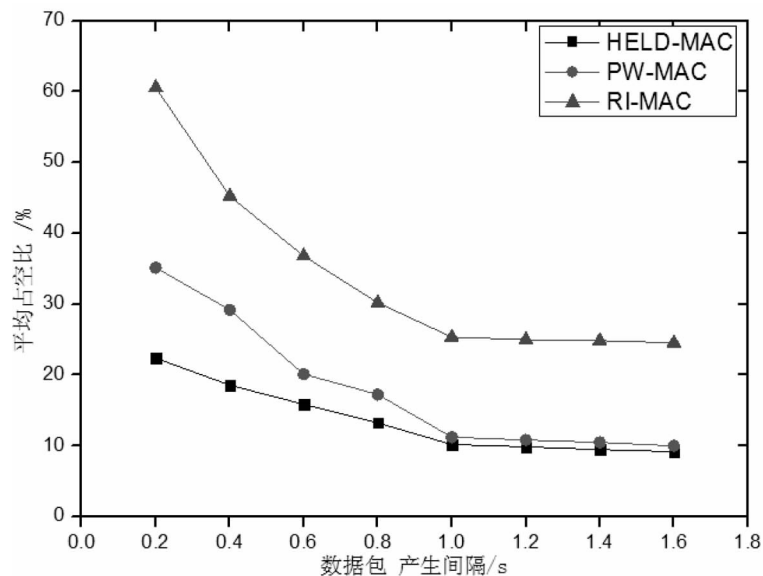


图 2-2 平均占空比比较

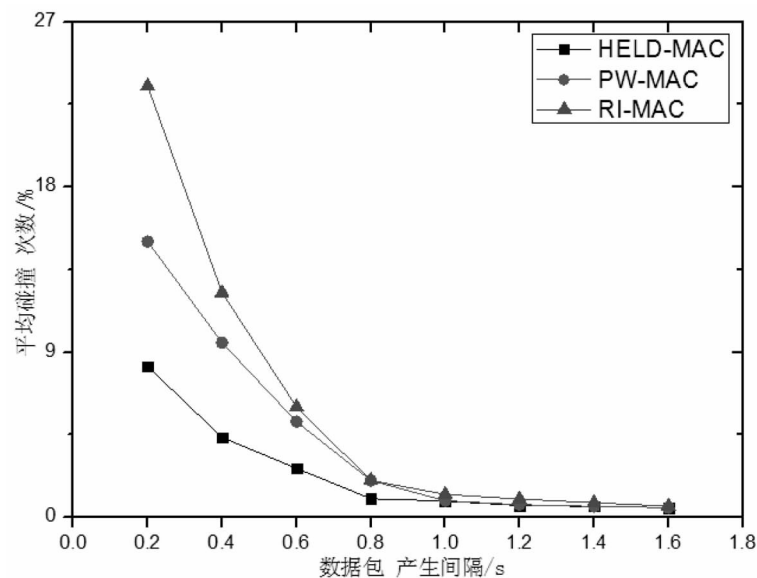


图 2-3 平均碰撞次数

图 2-2 和图 2-3 是三种协议在网格网络环境中平均占空比和平均碰撞次数对比图。随着数据包产生间隔逐渐减小,需要等待传输的数据包数量逐渐增加,由于 HELD-MAC 能够自适应改变节点的唤醒间隔,由图可看出,HELD-MAC 中数据传输碰撞次数相比于 RI-MAC 和 PW-MAC 增加趋势缓慢。因此 HELD-MAC 中节点数据重传的次数要比其它两种协议要少,发送节点的平均占空比要比 RI-MAC 和 PW-MAC 有较大优势。

2) 高速率、低延迟性能对比

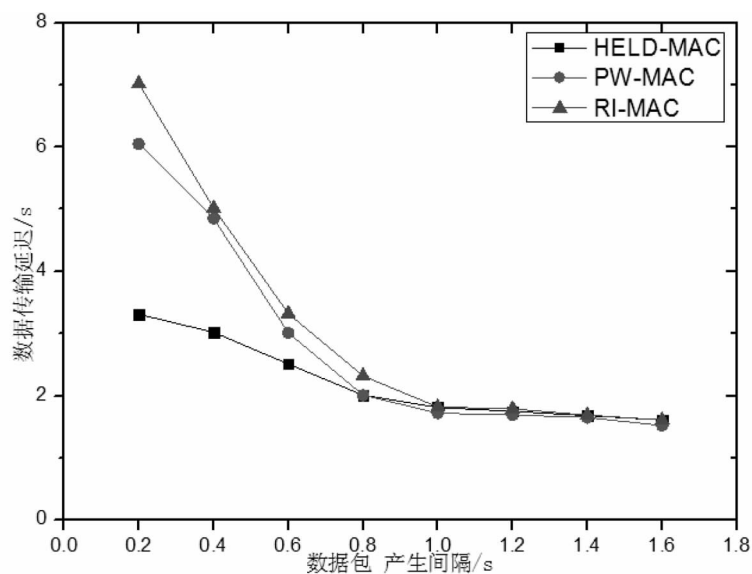


图 2-4 平均传输延迟

图 2-4 是三种协议在网格环境中数据平均传输延迟对比图。HELD-MAC 中接收节点可以根据发送节点负载情况自适应地改变节点的唤醒间隔,当负载增大时,节点减小唤醒时间间隔,当负载减小时,增大节点的唤醒时间间隔。因此随着数据产生时间间隔减小,当数据传输冲突时,数据重传次数要比其它两种协议要少,数据传输延迟要小。

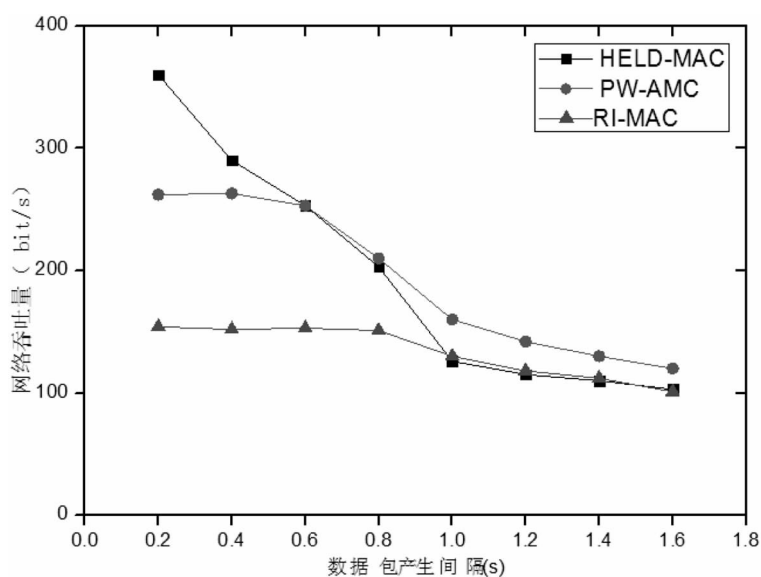


图 2-5 网络吞吐量比较

图 2-5 是三种协议在网格环境中网络吞吐量对比图。在时间间隔为 8ms 之前,网络负载不是很大使得 HELD-MAC 协议中节点的唤醒时间间隔为最大值,因此整个网络吞吐量要比其它两种协议

要小。随着网络负载的逐渐增加,HELD-MAC 协议优势逐渐体现,接收节点唤醒间隔减小,数据传输延迟减小,使得网络吞吐量较其它两种协议要大。

总体来讲,随着网络负载的逐渐增加,HELD-MAC 在平均占空比、数据平均传输延迟、数据传输碰撞次数方面呈线性增加,稳定性较强,因此 HELD-MAC 协议在低功耗和高速率方面相比其它协议具有明显的优势,更加适合传输大数量且波动较大的应用。

参 考 文 献

- [1] 孙利民,陈渝,朱洪松. 无线传感器网络 [M]. 北京:清华大学出版社,2005
- [2] 张园园. 中高速传感器网络高能效异步竞争型 MAC 协议研究与设计[D]. 广东工业大学硕士论文,2013
- [3] Akyildiz I F, Melodia T, Chowdhury K R. A survey on wireless multimedia sensor networks[J]. Computer networks, 2007, 51(4): 921-960
- [4] Kandis D, Tsagkaropoulos M, Politis I. Energy efficient and perceived QoS aware video routing over wireless multimedia sensor networks[J]. Ad Hoc Networks, 2011, 9(4): 591-607
- [5] Mark Stemm and Randy H Katz. Measuring and reducing energy consumption of network interfaces in hand-held devices[J]. IEICE Transactions on Communications, 1997, E80(8): 1125-1131
- [6] Oliver Kasten, Energy Consumption. <http://www.inf.ethz.ch/~kasten/research/bathtub/energy-consumption.html>
- [7] S. Ergen, P. Varaiya. PEDAMACS: power efficient and delay aware medium access protocol for sensor networks[J]. IEEE Transactions on Mobile Computing, 2006, 5 (7): 920 - 930
- [8] L. Van Hoesel, P. Havinga. A Lightweight medium access protocol(LMAC) for wireless sensor networks[C]. Proceedings of the 1st International Workshop on Networked Sensing Systems (INSS), 2013
- [9] W. Ye, J. Heidemann, D. Estrin. Medium access control with coordinated adaptive sleeping for wireless sensor networks[J]. IEEE/ACM Transactions on Networking, 2012, 12 (3): 493 - 506
- [10] T. van Dam, K. Langendoen. An adaptive energy-efficient MAC protocol for wireless sensor networks[C]. Proceedings of the First International Conference on Embedded Networked Sensor Systems (SenSys '03), 2003: 171
- [11] W. Ye, J. Heidemann, D. Estrin. An energy-efficient MAC protocol for wireless sensor networks[C]. Proceedings of the Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies, 2002, 3: 1567 - 1576
- [12] J. Polastre, J. Hill, D. Culler. Versatile low power media access for wireless sensor networks[C]. Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys '04), 2004
- [13] A. El-Hoiydi, J. Decotignie. WiseMAC: an ultra low power MAC protocol for multi-hop wireless sensor networks[J]. Algorithmic Aspects of Wireless Sensor Networks, 2004: 18 - 31
- [14] J. Polastre, J. Hill, D. Culler. Versatile low power media access for wireless sensor networks[C]. Proceedings of the 2nd International Conference on Embedded Networked Sensor Systems (SenSys '04), 2004
- [15] S. Lim, S. Kim, J. Cho, S. An. Medium access control with an energy-efficient algorithm for wireless sensor networks[J]. Proceedings of the Personal Wireless Communications, 2013: 334 - 343

High-speed and low-power wireless sensor network data transmission technology research and design

Gao Honglei

(School of Electronic and Information Engineering, North China Univ. of Tech., Beijing, 100144)

Abstract Wireless sensor networks generally used to transmitting information, such as temperature and humidity. When transferring images, audio and video and other multimedia data, not only the transfer rate gets attention, but also further highlights its node energy consumption and timeliness of information transmission problems. In Sensor network protocol stack, MAC layer directly controls the wireless RF module and the protocol design directly affects the energy consumption of nodes and data transmission timeliness, therefore it is an important part of the sensor network research.

This paper against the challenge that the traditional wireless sensor network are facing, when transfer large amounts multimedia information, based on the analysis of the MAC protocol, research is doing on the perspective of an asynchronous MAC protocol.

In wireless sensor networks, the existing receiver-initiated MAC protocol based on asynchronous duty cycling, are almost using pseudo-random wakeup schedule of neighbor nodes, in predicting the receiving node wakeup time. In the application of the transmission of multimedia information, load conditions in the network is dynamic, this method can't change the wakeup interval of node dynamically, and lead to higher data conflict rates while transmitting and larger data transmission delay. To solve this problem, the paper presents HELD—MAC (High energy efficiency and Low delay MAC), a new protocol that can ensure a different wakeup time between the nodes and predicate the wakeup time of receiving node accurately. The protocol can change the wakeup interval of node dynamically according to the load conditions. Meanwhile, in order to alleviate the problem of energy hole, the protocol can adaptively change the lowest node wakeup interval according to the residual energy of the node. Through simulation, HELD—MAC protocol has obvious advantages in the performance of data transmission delay, power consumption, networks throughput and transmission collision, compared with RI—MAC and PW—MAC protocols. The average data transmission delay and the average duty cycle of HELD—MAC is less than RI—MAC 81.2%, 19.5% respectively and less than PW—MAC 11.2%, 10.1%. HELD—MAC protocol is more suitable to large number of transmission and the transmission time-demanding wireless sensor network applications compared with others.

Key words wsn; media; MAC; asynchronous

面向 HEVC 的屏幕内容帧内编码相关技术的研究

研究生 张 扬

导师 张萌萌

(北方工业大学电子信息工程学院 北京 100144)

摘 要 随着云计算技术的迅猛发展,终端设备的图形化界面成为多媒体传输中不可或缺的重要内容,这直接促进了屏幕内容编码技术的发展。为了开发高效的屏幕内容编码技术,在制定了两版高效视频编码(High Efficiency Video Coding, HEVC)标准后,联合视频编码工作组(Joint Collaborative Team on Video Coding, JCT-VC)将屏幕内容编码标准作为重要的研究课题,致力于推出 HEVC 屏幕内容编码标准。

终端设备产生的屏幕内容视频既包括自然图像也含有文字、图表、小图标等人造内容,在采用帧内编码时往往会因为缺少有效的预测而难以获得高效的编码。同时,HEVC 复杂的编码单元和预测模式选择引入了较大的编码复杂度,使得屏幕内容编码难以满足应用对于实时性的要求。为此,本文重点研究 HEVC 标准在屏幕内容帧内编码的几项技术,针对已经推出的两版 HEVC 标准和正在研究的 HEVC 屏幕内容标准草案,从压缩效果和编码复杂度两个方面展开研究。主要研究成果包括:

1. 本文提出了改良的线状帧内块拷贝方案。该方案对 HEVC 屏幕内容编码标准草案中提出的帧内块拷贝技术进行优化改良,利用更优的线状块的替换矢量(Displacement Vectors, DV)预测和翻转拷贝方法来提高线状帧内块拷贝技术在屏幕内容帧内编码中的压缩效果。实验表明,对于屏幕内容帧内编码,在采用宽度为 1 的线条时,该方案平均能带来 11% 的 BD-rate 提升。

2. 本文提出了一种 HEVC 屏幕内容帧内编码的快速 CU 分割方案。该方案根据屏幕内容视频的统计规律,通过减少 CU 分割情况和编码工具的选择,在编码性能不产生明显变化的情况下大幅度减少编码时间,满足桌面共享、远程控制等应用对实时性的较高要求。对于 HEVC 屏幕内容编码标准,该方案能减少 29% 的屏幕内容帧内编码复杂度;对于第一版 HEVC,该方案能减少 38% 的屏幕内容帧内编码编码时间。

3. 本文中提出了一种屏幕内容帧内编码的去噪方案。针对前两版 HEVC 标准在低码率的屏幕内容帧内编码时存在失真较大的问题,本文提出的方案利用屏幕内容的重复性有效去除解码图像的量化噪声和去块滤波噪声,从而提高压缩图像质量。该方案应用于解码端,通过对解码图像进行块匹配和一维滤波,可以在不修改编码器的情况下提高屏幕内容帧内编码系统的性能。实验结果表明该方法在第二版 HEVC 的解码端能够平均提升 1.6dB 的 PSNR,并能大大改善文字区域的主观质量。

关 键 词 帧内块拷贝; 帧内编码; 屏幕内容编码; HEVC

作者简介 张杨,男,1988 年出生,硕士研究生,主要研究方向为视频编解码。

E-mail: zyjohnyoung@126.com

1 绪论

随着云计算技术的迅猛发展,单一的文本、语音和视频的网络通信已经不能满足人们的需求,人们需要传输终端设备的图形化界面(Graphic User Interface, GUI)来进行远程控制、监视、计算机支持的协同工作等。这些都离不开屏幕内容编码技术的支持。随着终端设备的发展和网络的普及,屏幕内容编码技术越来越多的应用于网络多媒体会议、远程办公、远程教学、产品演示、股票分析系统、以及网络

游戏等应用中。

这些利用屏幕内容编码技术的应用(如桌面共享,远程桌面等)不但可以在当前的数字设备上,与远程的计算机进行信息交互,还可以操控远程的计算机来解决当前资源无法解决的问题,实现广泛的资源(包括软件资源和硬件资源)共享^[1]。然而,在这些应用中,屏幕内容编码技术是关键性技术,往往决定着整个系统的性能。与传统自然图像/视频内容相比,屏幕内容视频序列(如:web 网页、PDF 文件、幻灯片文件、在线游戏等)还包含了由计算机产生的大量人造视觉内容。这些人造视觉内容主要是由文本、图表、小图标和自然图像等信息复合而成,是一种复合内容的视频序列。

目前国际上的图像/视频编码标准,如 JPEG 系列、H. 26X 系列、MPEG-X 系列^[2]及最新兴起的高效视频编码(High Efficiency Video Coding, HEVC)标准^[3]等,都是针对传统的自然图像/视频内容而制定的。然而,具有广泛应用领域的屏幕内容图像/视频与传统自然图像/视频有着本质的区别。屏幕内容没有自然图像/视频内容那样光滑,具有截然不同的信号统计特性。尤其是复合屏幕内容中的文本和图形内容,往往有着较为尖锐的边缘特征和较高的对比度,其中轻微的瑕疵也极易被用户察觉。因此,如果利用现有图像/视频编码标准压缩屏幕内容,极有可能对其中的高频信息造成损失,从而严重影响解码的视觉质量。在 HEVC 屏幕内容编码中,帧内编码由于不能利用帧间的相关性,要得到高质量的压缩效果,是一个较大的挑战。

为此,本课题将在最新 HEVC 标准框架的基础上,研究充分利用 HEVC 标准特性的屏幕视频帧内编码算法。主要研究内容包括:结合 HEVC 屏幕内容编码的帧内块拷贝工具,研究改良的线状块拷贝方案;利用 HEVC 标准中复杂的编码单元划分和屏幕内容视频统计规律,研究 HEVC 屏幕内容帧内编码的快速 CU 分割算法;针对屏幕内容的特点和 HEVC 帧内编码的量化噪声特点,研究针对屏幕内容编码的图像去噪方案。

目前,HEVC 的第二版视频编码标准已经尘埃落定,但国际上针对 HEVC 标准框架下的屏幕编码研究却是方兴未艾。本项目的研究成果可以进一步应用到高清的视频通信、视频监控、视频会议等领域,对于探索新的视频编码技术、解决高质量的屏幕内容在云计算环境下的传输等问题具有重要的理论意义和应用价值。本项目的研究成果也将进一步推动我国视频技术的发展,为云计算环境下的新一代视频编码标准做出贡献,前景广阔。

2 正文

2.1 HEVC 标准及屏幕编码简介

HEVC 是由隶属于国际电信联盟远程通信标准化组织(ITU-T)的视频编码专家组和隶属于国际标准化组织、国际电工委员会(ISO/IEC)的运动图像专家组联合组成的视频编码联合组(Joint Collaborative Team on Video Coding, JCT-VC)制定的新一代视频编码标准^[3]。

2013 年 1 月 HEVC 标准得到 ITU-T 的初步批准,作为第一版 HEVC 标准。之后第一版的 HEVC 标准被敲定为 ITU-T H. 265 和 ISO/IEC 23008-2。

2014 年 10 月份推出了 HEVC 范围扩展标准也就是 HEVC 的第二版标准。

值得一提的是在范围扩展标准的研究过程中,屏幕内容编码 HEVC 范围扩展标准中一个重要课题,人们提出了许多针对屏幕内容编码的改进。随着研究的深入,屏幕内容编码作为一个单独的课题,从范围扩展标准的研究中脱离出来,在第二版 HEVC 标准推出后依然还在研究中。

2.2 HEVC 屏幕内容帧内编码的改良线状帧内块拷贝方案

与^[4]中提到的 line-based IntraBC 一样,在本方案中当前的 $2N \times 2N$ 尺寸的 CU 被均匀的分成 $M \times 2N$ 或者 $2N \times M$ 大小的 line。用一个 flag 来区分两种尺寸的分割方式。每一 line 通过估计 DV 指向的重构区域的预测准确度和传输 DV 的开销得到的最优 DV。由于在一个 CU 中会有 $2N/M$ 个 line,需要消耗大量 bit 去编码 $2N/M$ 个 DV。另一方面,为了降低复杂度,最优的 DV 只能在特定的区域搜索,有可能找不到最佳匹配的 line。为了处理这些问题,我们分别引入了 DV 预测和翻转块拷贝预测来减少 DV 的开销并提高预测准确度。

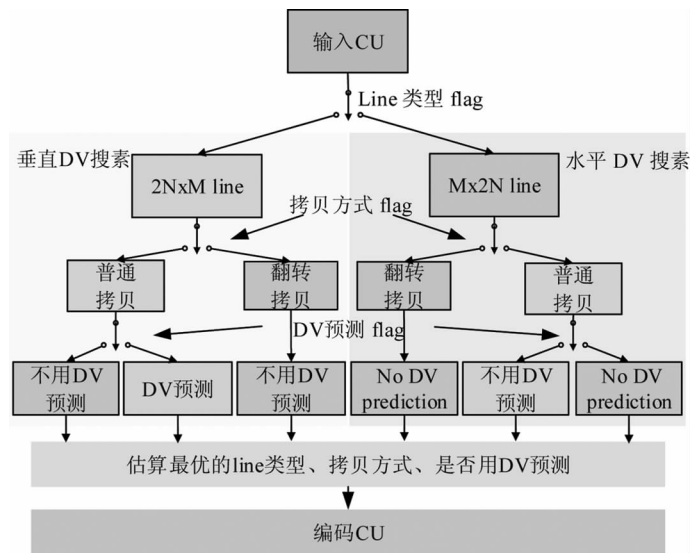


图 1 编码方案流程

加入 DV 预测和翻转块拷贝的线状 IntraBC 的流程如图 1 所示。这里用了三个 flag 来区分 line 的类型、是否用 DV 预测,以及块拷贝方式。编码端通过计算所有可能的 $COST_{DV}$ 来估计一个最优编码方式。为了减少编码复杂度,线状 IntraBC 跳过大于 16×16 的 CU。

编码方案在最新的 HEVC 参考程序 HM-12.1+RExt-5.1 上实现,采用 JCT-VC 推荐的 all-intra lossy 屏幕内容编码环境。测试序列包括 YCbCr 4:4:4 SC 和 YCbCr 4:4:4 (Optional) 两组。

表 1 没有 DV 预测和翻转块拷贝的线状 IntraBC

组	序列	Y BD-rate(%)		
		M=1	M=2	M=4
YCbCr 4:4:4SC	sc_map	-2.0	-2.2	-2.0
	sc_web_browsing	-8.7	-13.9	-13.0
	sc_wordEditing	-6.2	-8.0	-8.2
	sc_SlideShow	-2.1	-1.7	-1.6
	sc_programming	-4.3	-4.8	-4.5
	sc_ppt_doc_xls	-13.7	-12.0	-8.0

组	序列	Y BD-rate(%)		
		M=1	M=2	M=4
YCbCr 4:4:4SC (Optional)	sc_cad_waveform	-18.1	-18.8	-15.0
	sc_pcb_layout	-14.6	-11.0	-7.6
	sc_cg_twist_tunnel	-3.4	-3.5	-3.3
	平均	-8.1	-8.4	-7.0
编码时间增加(%)		23	23	21
解码时间增加(%)		-3	-10	-12

表 1-3 显示了本文提出技术的实验结果。表 1 展示没有用 DV 预测和翻转块拷贝的线状 Intra-BC 的结果。表 2 展示了增加了 DV 预测的线状 IntraBC 的结果。可以看到,当 $M=1$ 时,DV 预测平均节省了 2.9% 的 bits,跟表 1 比编码时间几乎没有变化。

加入 DV 预测和翻转块拷贝的结果在表 3 显示,由于翻转块拷贝在 $M=1$ 时不使用,所以表 3 只有 M 等于 2 或者是 4 的结果。可以看到,当 $M=4$ 时,该方案比表 2 的结果平均节省 1.8% 的 bit。对于序列“sc_cad_waveform”通过更好的处理对称内容,节省的 bit 更是高达 6.7%。

表 2 有 DV 预测的线状 IntraBC

组	序列	Y BD-rate(%)		
		M=1	M=2	M=4
YCbCr 4:4:4SC	sc_map	-2.6	-2.3	-2.0
	sc_web_browsing	-13.1	-16.1	-14.0
	sc_wordEditing	-9.7	-9.4	-8.8
	sc_SlideShow	-3.9	-2.3	-1.7
	sc_programming	-6.2	-5.5	-4.7
	sc_ppt_doc_xls	-15.5	-12.2	-8.4
YCbCr 4:4:4SC (Optional)	sc_cad_waveform	-22.7	-19.5	-15.5
	sc_pcb_layout	-19.9	-11.6	-8.1
	sc_cg_twist_tunnel	-5.7	-4.5	-3.8
	平均	-11.0	-9.3	-7.5
编码时间增加(%)		22	22	20
解码时间增加(%)		-5	-7	-8

表 3 有 DV 预测和翻转块拷贝的线状 IntraBC

组	序列	Y BD-rate(%)	
		M=2	M=4
YCbCr 4:4:4SC	sc_map	-2.6	-2.5
	sc_web_browsing	-14.7	-14.0
	sc_wordEditing	-10.0	-10.2
	sc_SlideShow	-3.3	-3.4
	sc_programming	-6.3	-5.9
	sc_ppt_doc_xls	-14.0	-10.7
YCbCr 4:4:4SC (Optional)	sc_cad_waveform	-20.5	-17.6
	sc_pcb_layout	-16.4	-14.8
	sc_cg_twist_tunnel	-5.4	-5.0
	平均	-10.3	-9.3
编码时间增加(%)	28	27	
解码时间增加(%)	-5	-5	

本文提出一个改进版的线状 IntraBC 方案来处理屏幕内容编码。该方案并没有显著增加复杂度。它通过 DV 预测和翻转块拷贝预测来分别减少 DV 开销,增加预测精度。跟 HEVC 范围扩展标准相比,该方案能显著改进编码效果。

2.3 HEVC 屏幕内容帧内编码的快速 CU 分割方案

在本文的方案中,CU 区域仍然被分成三类:平滑块、普通块和尖锐块。寻找平滑块和尖锐块的流程如图 2 所示。一个编码树单元首先被分成 16 个 16×16 的区域。在每个 16×16 的方块里,每一个像素都需要查看是否平滑,如果查找到一个孤立像素就停止,剩下的就不用查看了。如果在一个方块里所有的像素都是平滑像素,这个方块就被认作是一个深度为 2 的平滑像素。如果其中 32×32 的 CU 区域有 4 个 16×16 的平滑块,这个 32×32 的 CU 区域就会被认作也是一个平滑块。深度为 0 的平滑块即是拥有 4 个 32×32 的平滑块的 CTU。

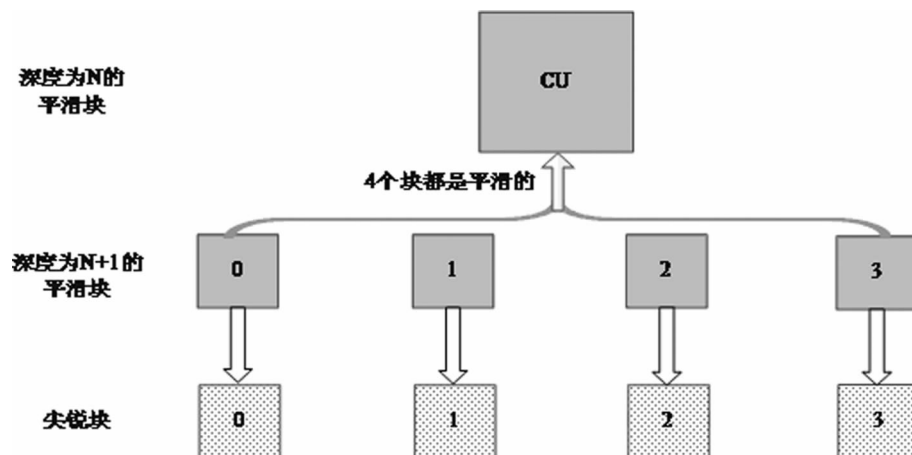


图2 寻找平滑块和尖锐块的流程图

快速方案如图3所示。 64×64 的普通块,总是会分割成更小的CU,但是是否跳过RD检查需要看周围CU的深度值,这是因为在最优CU分割中当前CU的深度跟周围CU的深度有很强的相关性^[5]。结果如表4所示。

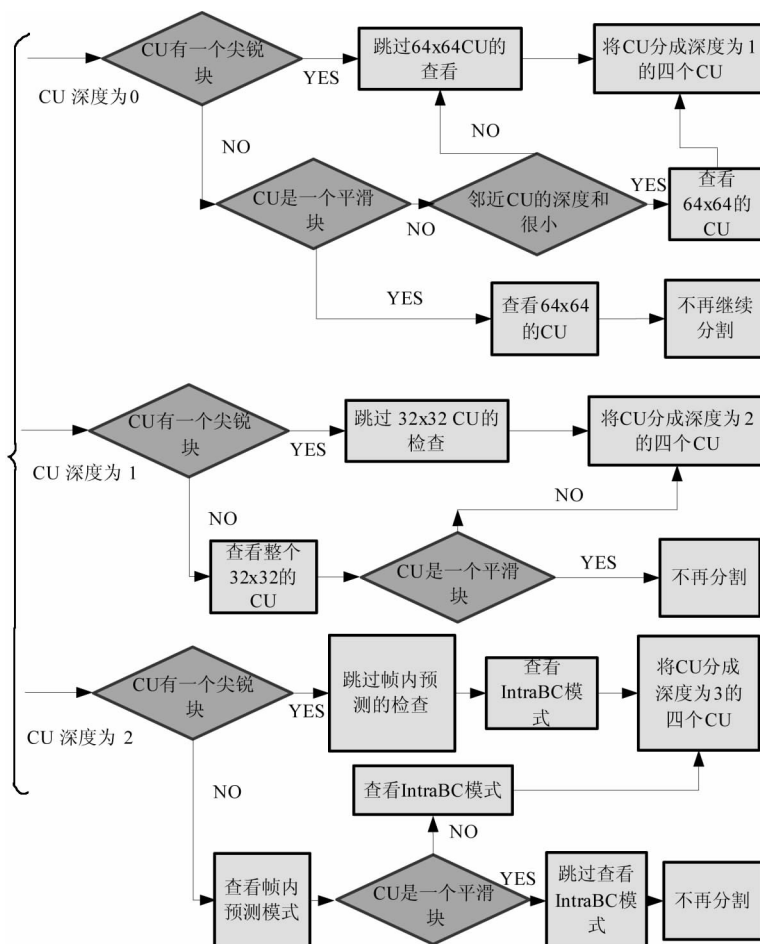


图3 快速算法流程图

表 4 屏幕内容编码标准草案下的屏幕内容视频测试结果

组	视频序列	BD-rate(%)			$\Delta T(\%)$
		Y	U	V	
YCbCr 4:4:4SC	sc_map	0.0	0.0	0.4	19
	sc_web_browsing	0.4	0.4	0.3	25
	sc_wordEditing	0.5	0.3	0.4	31
	sc_SlideShow	0.6	0.6	0.7	49
	sc_programming	0.7	0.8	0.9	20
	sc_ppt_doc_xls	0.6	0.8	0.6	25
YCbCr 4:4:4 SC (Optional)	sc_cad_waveform	1.9	2.0	2.0	24
	sc_pcb_layout	2.6	2.7	2.6	29
	sc_cg_twist_tunnel	0.8	0.9	0.8	38
平均		0.9	0.9	1.0	29

实验结果可以看出,在 HEVC 扩展标准中,本文提出的快速方案可以在 BD-rate 只增加 0.9% 的情况下减少大约 29% 的编码时间。

尽管 HEVC 扩展标准可以有效地提升屏幕内容视频的编码效果,但是由于其还处于研究阶段,大量的应用还是基于默认的 HEVC 标准。

把快速方案关于 IntraBC 部分的内容去掉,这个快速方案也可以应用于默认的 HEVC 标准,而且还能获得更多的时间节省。屏幕内容实验结果如表 5 所示。由于第一版 HEVC 只支持 4:2:0 的采样格式,所以所有的 4:4:4 的视频序列都做了 4:2:0 的采样使之能够支持第一版 HEVC 的参考模型。

表 5 第一版 HEVC 标准在屏幕内容视频的测试结果

组	视频序列	BD-rate(%)			$\Delta T(\%)$
		Y	U	V	
YCbCr 4:4:4SC	sc_map	0.0	0.0	0.0	24
	sc_web_browsing	0.2	0.0	0.0	37
	sc_wordEditing	0.1	0.3	0.1	37
	sc_SlideShow	0.2	0.3	0.5	61
	sc_programming	0.1	0.5	0.5	40
	sc_ppt_doc_xls	0.1	-0.1	0.1	36
YCbCr 4:4:4 SC (Optional)	sc_cad_waveform	0.4	0.6	0.6	23
	sc_pcb_layout	0.2	0.4	0.4	40
	sc_cg_twist_tunnel	0.1	0.2	0.2	41
平均		0.1	0.2	0.3	38

2.4 屏幕内容帧内编码的去噪方案

由于屏幕内容图像中会有大量的重复内容,而这些内容往往是编码失真较大的文字。相同的文字

在原始视频中是完全相同的,而在编码后它们分别引入了量化噪声和去块滤波噪声。本文提出去噪方案找重复的内容,利用它们在原始视频中完全相同的特点,相互预测,去掉噪声。流程如图 4 所示。

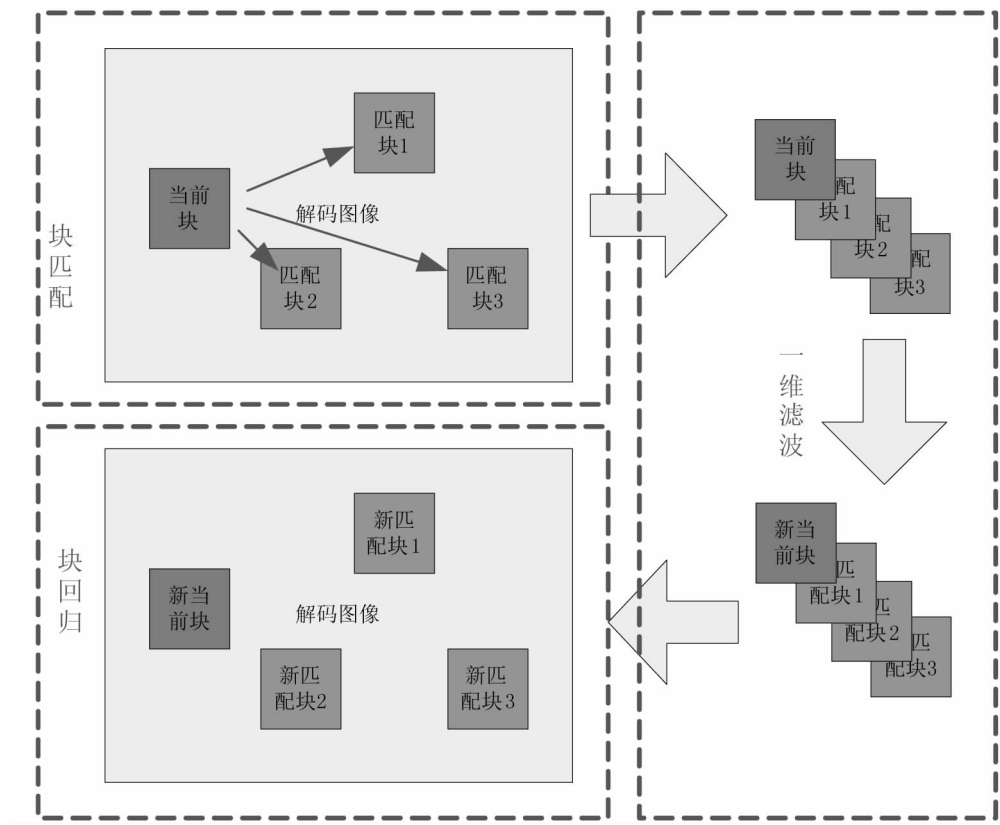


图 4 图像去噪方案中块的处理流程

本方案的实验是在屏幕内容序列解码端进行的。噪声图像是使用第二版 HEVC 参考软件 HM16.0 在 QP37 时采用帧内编码得到的。实验结果如表 7 所示。

表 7 图像去噪方案的实验结果

序列	解码后 PSNR(DB)	处理后 PSNR(DB)	PSNR 提升(DB)
programming	34.630451	35.965721	1.33527
slideediting	32.942103	34.399925	1.457822
SlideShow	36.064960	36.536226	0.471266
TwistTunnel	34.965341	35.876095	0.910754
VideoConferencingDocSharing	33.534618	35.196082	1.661464
webbrowsing	36.077054	38.169630	2.092576
WordEditing	34.225721	37.585950	3.360229

可以看出,本算法能有效的提升解码后图像的升编码质量,平均 PSNR 能有 1.6db 的提升。编码

后的文字的主观质量也有所改善,如图 5 所示,消除了明显的失真。



(a)解码图像



(b)去噪后图像

图 5 算法后的数据对比

2.5 结论

针对尚处于研究的 HEVC 屏幕内容编码标准,本文提出的改进的线状帧内块拷贝方案能够有效的提升编码的压缩性能,能用更少的带宽来获得更好的屏幕内容视频。针对屏幕内容编码,本文提出的 HEVC 屏幕内容帧内编码的快速 CU 分割方案,能在编码效果几乎没有损失的情况下大量的减少编码时间,提高了屏幕内容编码系统的实时性。对于未使用屏幕内容编码标准所编码的屏幕内容视频序列,本文提出的屏幕内容帧内编码去噪方案能在解码端有效去除解码图像中掺杂的量化和滤波噪声。

实验结果表明,本文对面向 HEVC 的屏幕内容编码系统的几个关键问题做了较好的改进,有效提升了屏幕内容编码系统的性能。

参 考 文 献

- [1] 吴婧. 桌面图像序列编码方法的研究. 浙江:浙江大学,2009
- [2] 高文,赵德斌,马思伟. 数字视频编码技术原理. 北京:科学出版社,2010
- [3] Sullivan, G. J. Overview of the High Efficiency Video Coding (HEVC) Standard. IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, 2012, 22(12): 1649–1668
- [4] C. C. Chen, T. S. Chang, R. L. Liao, W. H. Peng, H. M. Hang, C. L. Lin and F. D. Jou, “AHG8: Line-based Intra Block Copy,” JCTVC–O0205, 2013
- [5] L. Shen, Z. Zhang, and P. An, “Fast CU size decision and mode decision algorithm for HEVC intra coding,” IEEE Trans. Consumer Electron., vol. 59, no. 1, pp. 207–213, Feb. 2013

Research on Screen Content Intra Coding for HEVC standard

Zhang Yang

(School of Electronic and Information Engineering, North China Univ. of Tech. , Beijing, 100144)

Abstract With the rapid development of cloud computing technology, graphical user interface from terminal has become indispensable and important content for multimedia transmission, which improves the development of screen content coding technique. For a better development of screen content coding technology, JCT—VC (Joint Collaborative Team on Video Coding) has been trying to develop the HEVC screen content coding standard after two editions of HEVC released.

Screen content videos generated by the terminal device contain natural images, as well as texts, charts, small icons and other artifacts. It often cannot obtain high coding efficiency in intra-frame coding because of the lack of efficient prediction. At the same time, coding unit partition and mode decision in HEVC introduces great complexity, which is difficult to meet the need of real-time screen content coding applications. Therefore, this paper focuses on several technologies of the screen content intra coding for the both two released HEVC standard and the draft of HEVC screen content coding standard. The research includes coding efficiency and complexity. The main research achievement contains:

1. This paper presents a modified line-based intra block copy scheme. To achieve better screen content coding efficiency, the scheme refines the intra block copy in the draft of HEVC screen content coding standard. The scheme uses a better displacement vectors (DV) prediction and flipping linear block copy method to improve intra block copy. Experimental results show that the modified line-based intra block copy scheme brings an average of 11% of BD—rate increase when using a line with width of 1.

2. This paper presents a fast CU splitting scheme for HEVC screen contents intra coding. The scheme greatly reduces coding time with almost unchanged coding efficiency by reducing the choices of some CU partition and some coding tools based on statistic of screen content coding, in order to meet the real-time need of the applications using content coding technology like desktop sharing and remote control. For HEVC screen content coding standard, the program can reduce 29% complexity of the screen content intra coding; for the first edition of HEVC, the program can reduce 38% of the coding time for screen content coding.

3. This paper presents a screen content intra coding denoising scheme. As the first two versions of HEVC standard have great distortion in low bit rate, the proposed scheme, using repeatability of screen contents, can effectively remove the quantization noise and filtering noise in the decoded images, so as to improve the quality of decoded images. This scheme is applied to the decoder, and it can improve the performance of the screen content frame coding system without changing the encoder by

block matching and 1-D filtering. The experimental results show that the scheme can improve average 1.6 DB PSNR in the decoder of the second edition of HEVC, and greatly improve the subjective quality in text area.

Key words intra block copy; intra coding; screen content coding; HEVC

高精度北斗天线设计

研究生 石璐瑶

导师 宋 鹏 王恩成

(北方工业大学电子信息工程学院 北京 100144)

摘 要 北斗卫星导航系统是我国独立研制、兼容开放的全球卫星导航系统。随着北斗卫星导航系统的建设与完善,系统正朝着高精度化的方向发展。北斗天线作为导航系统的重要组成,其性能将直接决定北斗卫星导航系统测量定位的精度。天线相位中心变化和多径干扰是高精度测量的两个主要的误差源,同时也是高精度卫星导航天线设计的热点和难点。本文以高精度北斗天线设计入手,深入研究了宽带圆极化电磁偶极子天线的设计、抑制多径干扰的天线技术和建立天线相位中心模型的方法。论文主要研究内容为以下三个方面:

1. 研究了新型的互补天线单元—电磁偶极子天线,并将该研究应用到北斗天线的设计中,提出了一种新型的宽带圆极化电磁偶极子天线。所提出的天线主要由交叉电磁偶极子,双层反射板和左右手复合结构(CRLH)功分移相器组成。偶极子的上半部分是电偶极子,它们的一端分别与四个垂直的短路贴片相连,垂直短路贴片在此处等效为磁偶极子。实物测试结果表明该天线的工作频带宽,能够覆盖“北斗二号”系统的所有频段,而且在整个工作频带内,天线的辐射方向图性能好,相位中心稳定。该天线优异的性能使其完全符合高精度北斗天线的设计要求。

2. 研究了抑制多路径干扰的天线技术。着重研究了扼流圈技术,对扼流圈抑制多路径干扰的原理进行了研究。设计了用于北斗天线的扼流圈,并将所设计的高精度北斗天线安置在扼流圈上。结果表明扼流圈改善了天线的低仰角轴比性能,降低的天线的后瓣辐射,从而增强了天线抑制多路径干扰的能力。

3. 研究了建立天线相位中心模型的方法。参考国外高精度 GPS 天线的相位中心建模方法,根据国际全球卫星定位系统服务组织(IGS)的相位中心模型标准,建立了高精度北斗天线的相位中心模型。

关 键 词 电磁偶极子天线;圆极化;多路径;天线相位中心;北斗卫星导航系统

作者简介 石璐瑶,男,1989 年出生,硕士研究生,主要研究方向为射频电路与天线设计。

E-mail:you. are. too@163. com

1 绪论

全球卫星导航系统(Global Navigation Satellite System,GNSS)包括我国正在实施的自主研发并独立运行的北斗卫星导航系统(BeiDou Navigation satellite system,BDS),美国的全球定位系统(Global Positioning System,GPS)、俄罗斯的格洛纳斯(Global Navigation Satellite System,GLO-NASS)和欧盟的伽利略系统(GALILEO)^[1]。北斗系统的建设包含两个阶段,其中“北斗一号”系统于 2003 年开始建设,目前该系统已经建设完成并成功应用于测绘、交通和抗灾救灾等诸多领域,取得显著的社会效益。“北斗一号”系统是局域性的卫星导航系统,只能为我国和周边邻国提供服务,为此我国正在紧密地建设“北斗二号”系统,把北斗系统逐步建设成全球卫星导航系统^[2]。“北斗二号”系统工作在 B1 (1559. 052—1591. 788 MHz)、B2(1162. 220—1217. 370 MHz)和 B3(1250. 618—1286. 423 MHz)三个频段,系统将于 2020 年具备覆盖全球的服务能力,届时将对于国民经济发展和国防建设产

生重要的意义。

对于高精度北斗天线的设计,多路径干扰和天线相位中心变化是两个不可忽视的因素。卫星信号通过高山森林、海洋湖泊或者高耸建筑等环境时,会发生散射、反射产生多路径信号,这些多路径信号会和直达信号一起进入接收机,但相比直达信号,它们的相位和幅度产生了变化,这会降低接收机的测量精度,造成多路径干扰。为了抑制多路径干扰,需要从天线设计上采取措施,故此抑制多路径成为高精度北斗天线设计中的重要问题。高精度天线的设计另外一个重要问题是天线相位中心变化。在卫星导航测量中是以天线的几何中心为基准,但实际测量却是基于天线相位中心进行的,只有天线的几何中心与天线相位中心重合时,测量值才是真实值。而天线相位中心是天线的等效辐射中心,它对于理想天线是唯一的,但是实际中它在某一范围内变化,与入射信号方位角、俯仰角、信号频率和天线形式有关^[3]。因此,设计相位中心稳定的天线和修正天线相位中心变化是高精度北斗天线设计亟需解决的关键问题。

本课题研究了高精度北斗天线设计,对天线设计中的宽频带、频率稳定性、辐射方向图、抗多径能力和天线相位中心建模等关键环节进行深入的探究,提出一种高精度北斗天线的设计思路。本课题对北斗系统的建设和完善及其高精度的应用研究具有重要意义。

2 正文

2.1 高精度北斗电磁偶极子天线设计

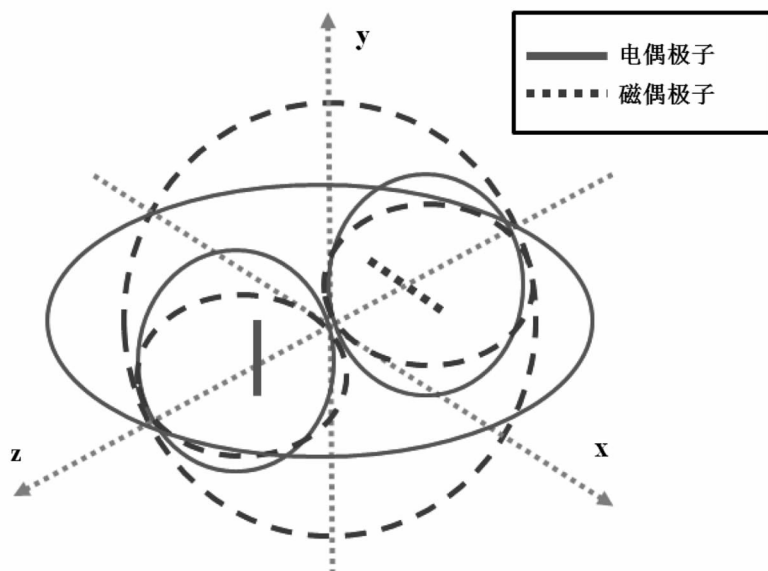


图 2.1 电偶极子与磁偶极子的方向图

图 2.1 展示了电偶极子和磁偶极子的方向图。其中实线表示电偶极子的方向图,虚线代表磁偶极子的。从上图可知,电偶极子的 E 面方向图类似数字“8”,H 面方向图类似字母“O”;相反的,磁偶极子的 E 面方向图为字母“O”,H 面为数字“8”。两个偶极子的方向图互补,如果将两个偶极子结合在一起,可以让天线的两个极化面一致,同时使辐射能量更加集中,在整个工作频段内可以获得更加一致的定向方向图。近来,Luk 等学者根据这一原理提出了一种成熟的宽带互补型电磁偶极子天线^[4],该电

磁偶极子天线单元具有阻抗带宽宽, E 面和 H 面方向图完全对称, 后瓣辐射小和天线性能稳定等优点。该天线性能优异, 众多学者在此基础上深入研究^[5-7]。交叉偶极子天线是一种常见的圆极化天线, 这种天线通过两个幅度相同相位相差 90° 的激励产生圆极化辐射。根据交叉偶极子和电磁偶极子的概念, 我们设计了这个新型的圆极化交叉电磁偶极子天线, 使其可以工作在北斗二代的全部频段, 而且天线在工作频段内性具有稳定良好的性能。在我们的设计中, 如图 2.2 所示, 天线主要有交叉电磁偶极子, 双层反射板和左右手复合结构 (CRLH) 移相器。偶极子的上半部分是电偶极子, 它们的一端分别与四个垂直的短路贴片相连, 在距离连接处 L_2 长度的地方, 电偶极子向下倾斜 θ 度角, 以展宽天线的波束, 减小天线的尺寸。电偶极子下倾部分的长度为 L_1 , 电偶极子宽 W 。与电偶极子相连的垂直短路贴片高为 H (约为中心频率的四分之一波长), 两个向对的垂直短路贴片的间距为 S , 垂直短路贴片在此处等效于磁偶极子。另外, 我们提出了应用于圆极化天线设计的窄边电磁偶极子模型。该模型的宽度 W 约为 0.06 个工作波长, 仅为宽带电磁偶极子天线的宽度的十分之一, 这种结构不仅缩小了天线尺寸, 而且便于天线的设计和优化。为了获得圆极化辐射和宽轴比带宽, 设计了具有宽频特性的 CRLH 移相器。CRLH 功分移相器的结构示意图如图 2.3 所示。

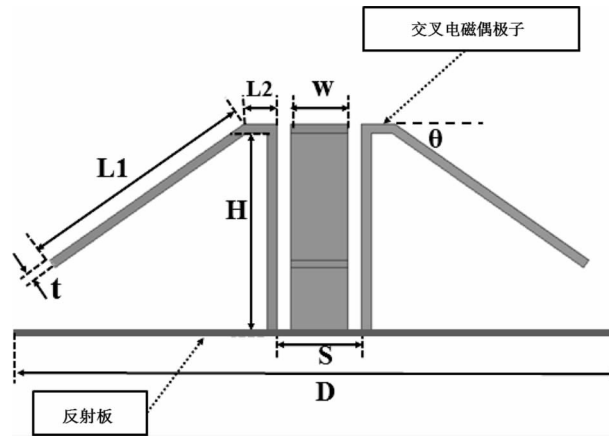


图 2.2 交叉电磁偶极子天线结构图

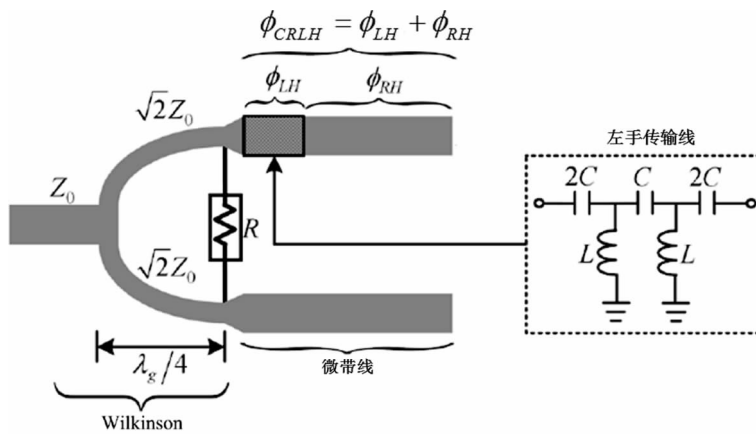
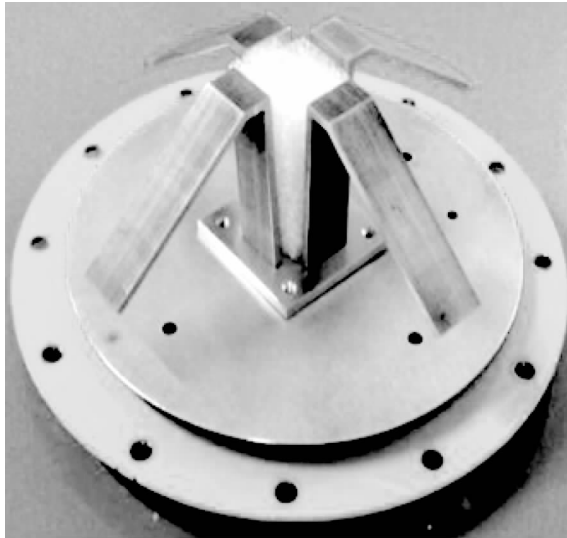


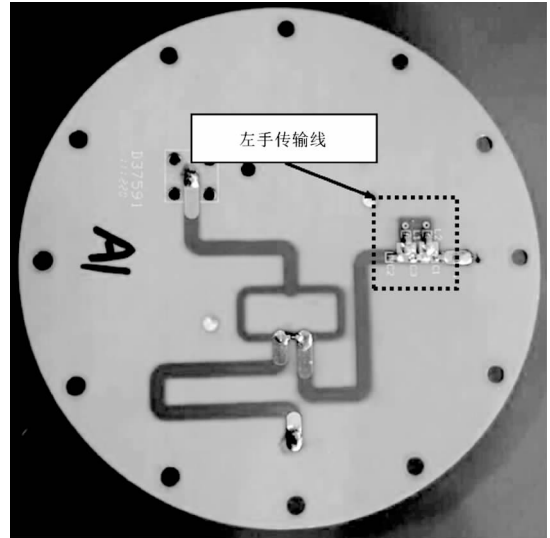
图 2.3 CRLH 功分移相器的结构示意图

为了验证仿真结果的正确性, 制作了天线的实物模型, 如图 2.4 所示。天线的偶极子由低损耗, 易

加工的铜板制成,反射板由 0.3mm 的紫铜皮制成。CRLH 功分移相器印制在 1.6 mm 的 FR4 介质基板上,FR4 基板的介电常数为 4.4。左手传输线由 0805 封装的贴片电感和电容实现。本天线使用普纳 PNA3623P 矢量网络分析仪测量了电压驻波比,使用微波暗室测量了天线的增益、轴比和天线方向图。



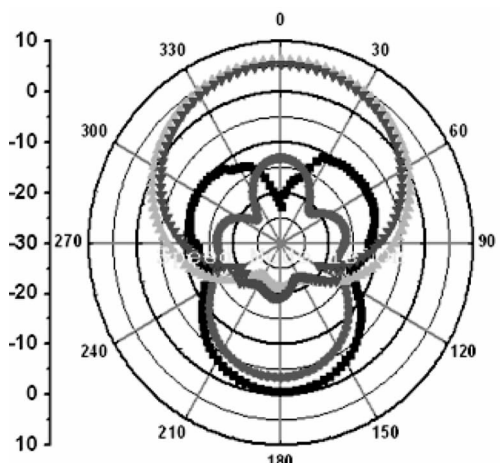
(a) 天线三维结构图



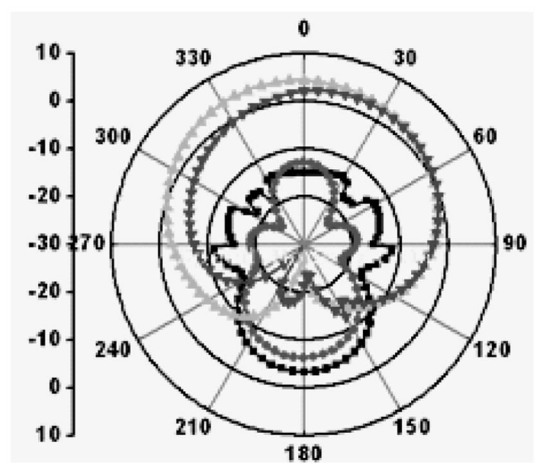
(b) CRLH 功分移相器图

图 2.4 天线实物模型图

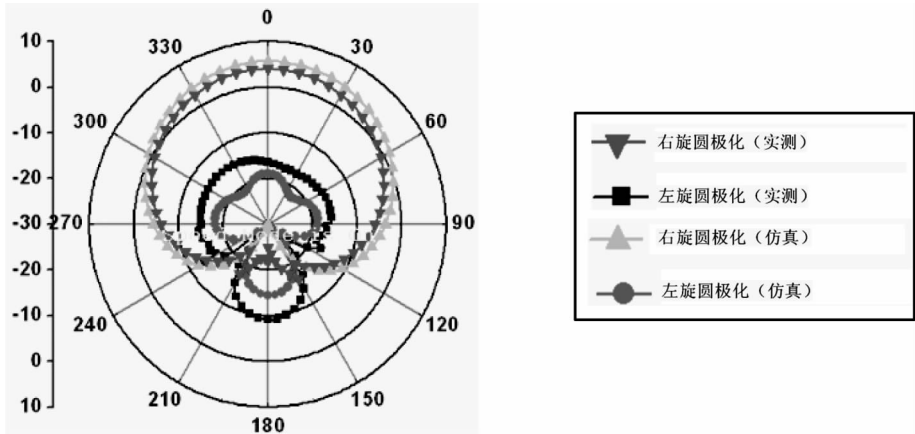
图 2.5 给出了天线辐射方向图,从图中可以看出,天线在工作频段具有对称,稳定的方向图,天线的 3dB 波瓣宽度较宽,达到了 90° 左右,并且天线的低仰角增益较高(大于 -5 dBi)。天线的电压驻波比和增益的仿真和测试结果如图 2.6 所示,该天线的阻抗带宽($VSWR \leq 1.5$)为 $1.1 \text{ GHz} \sim 1.85 \text{ GHz}$,相对带宽达到了 51.4%。在工作频段内,天线增益非常稳定,保持在 $5.6 \pm 0.6 \text{ dBi}$ 范围内。图 2.7 是天线轴比的测试和仿真结果,从图中可以看出天线的轴比带宽达到了 42.9%(1.1 GHz 到 1.7 GHz)。上述结果表明天线的仿真和测试结果非常吻合。从实验结果可以看出,将交叉偶极子原理与



(a) 1.207 GHz 处方向图



(b) 1.268 GHz 处方向图



(c) 1.561 GHz 处方向图

图 2.5 天线方向图的测试和仿真结果

电磁偶极子技术结合的方案,不仅提高了天线的匹配能力,而且该方案也极大改善了圆极化天线的方向图,使其符合高精度北斗天线方向图的特殊要求。

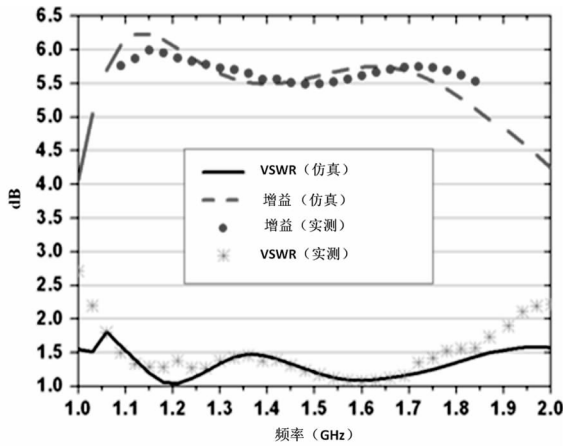


图 2.6 天线增益和 VSWR 的仿真测试结果

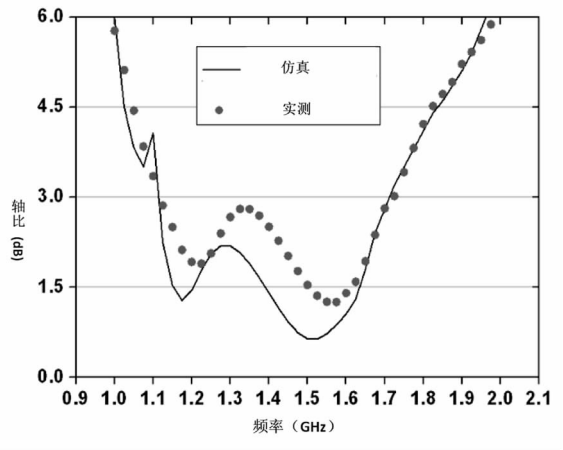
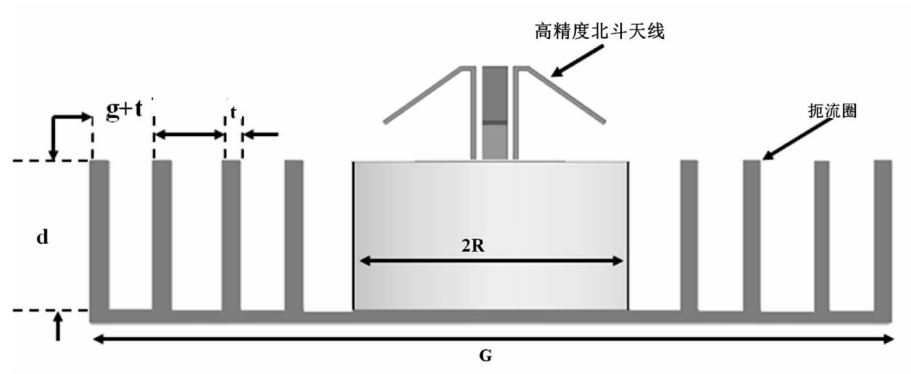


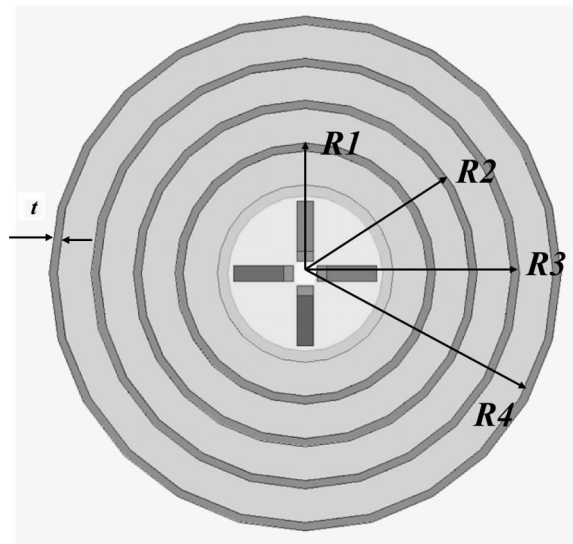
图 2.7 天线轴比仿真和实测结果

2.2 抑制多径干扰的扼流圈设计

基于电磁理论分析,扼流圈的工作原理可以用金属波纹表面理论解释^[8],对金属波纹表面的理论分析,当扼流圈槽深满足 $\lambda/4 \leq d \leq \lambda/2$ 的条件时,扼流圈可以有效地抑制表面波,从而降低后瓣辐射和交叉极化电平,提高天线抑制多路径干扰的能力。基于此原理,设计了与高精度北斗天线配套的扼流圈,其仿真结构图如图 2.8 所示。天线放置扼流圈上,与扼流圈一起在 HFSS 中仿真。通过观察天线方向图的变化,来优化扼流圈的尺寸。



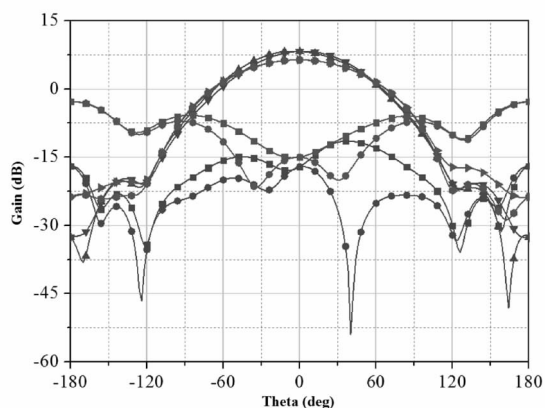
(a) 侧视图



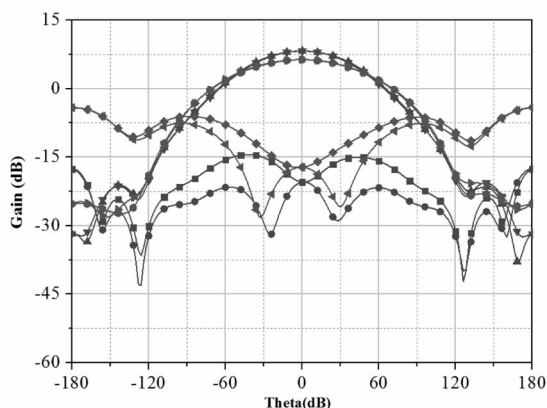
(b) 俯视图

图 2.8 扼流圈结构图

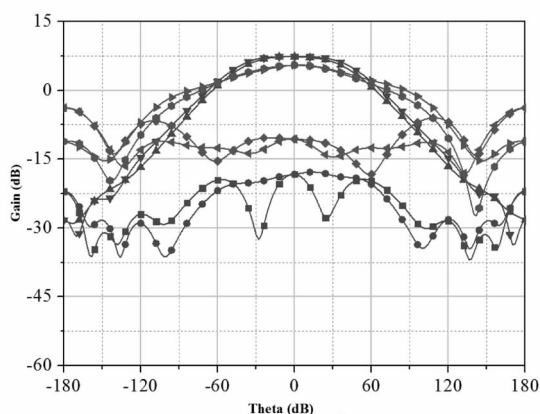
为了说明扼流圈起到抑制多路径干扰的作用,对加载扼流圈的电磁偶极子天线和不加载扼流圈的电磁偶极子天线分别仿真,通过比较仿真结果进行说明。图 2.9 分别比较了单独的电磁偶极子天线和加载扼流圈的电磁偶极子天线在 1.207 GHz、1.268 GHz 和 1.561 GHz 时 E/H 面方向图。从图中可以看出,扼流圈显著降低了三个频点的交叉极化电平和背瓣辐射,提高天线抑制多路径的能力。



(a) 1.207GHz 处方向图



(b) 1.268 GHz 处方向图



(c) 1.561 GHz 处方向图

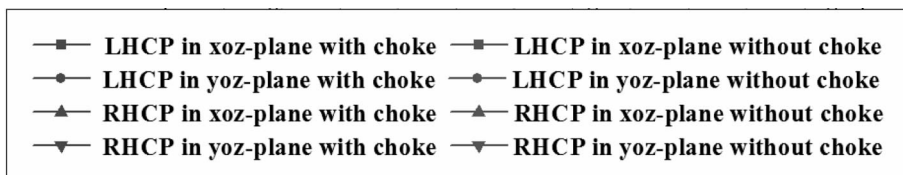
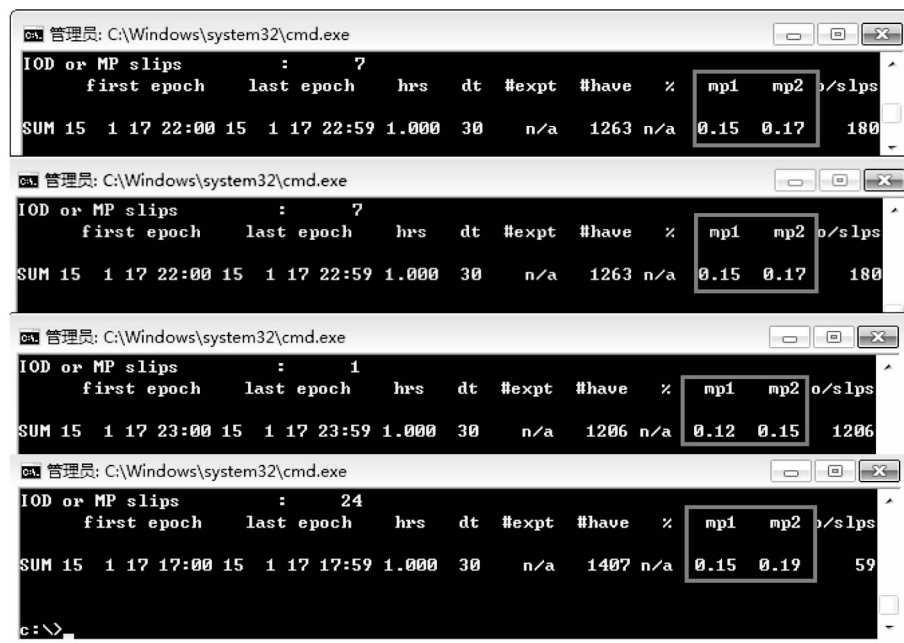


图 2.9 天线有无扼流圈时的辐射方向图

为了验证扼流圈设计的可行性,加工制作了如图 2.10 所示的扼流圈。为了更好的说明扼流圈对多径抑制的作用,将这个加载扼流圈的电磁偶极子天线应用到现场测试环境中,采用 TEQC 软件对扼流圈天线进行质量检核^[9],部分多路径检核结果见图 2.11 所示。从图中可以看出多径指标 MP1、MP2 的值均小于 0.2 m,远好于工程规定的 0.5 m。实测结果说明扼流圈天线有效地抑制了多路径干扰,提高了北斗卫星接收机的测量定位精度,验证了扼流圈设计的可行性。



图 2.10 扼流圈天线实物图



IOD or MP slips	:	7
first epoch	last epoch	hrs
SUM 15	1 17 22:00 15	1 17 22:59 1.000
dt	30	#expt
#have	1263	%
mp1	0.15	mp2
o/s lps	180	

IOD or MP slips	:	7
first epoch	last epoch	hrs
SUM 15	1 17 22:00 15	1 17 22:59 1.000
dt	30	#expt
#have	1263	%
mp1	0.15	mp2
o/s lps	180	

IOD or MP slips	:	1
first epoch	last epoch	hrs
SUM 15	1 17 23:00 15	1 17 23:59 1.000
dt	30	#expt
#have	1206	%
mp1	0.12	mp2
o/s lps	1206	

IOD or MP slips	:	24
first epoch	last epoch	hrs
SUM 15	1 17 17:00 15	1 17 17:59 1.000
dt	30	#expt
#have	1407	%
mp1	0.15	mp2
o/s lps	59	

图 2.11 TEQC 软件的多路径检核结果

2.3 天线相位中心模型建立

为了完整描述天线相位中心特性和方便建立天线相位中心数学模型,需要引入天线几何中心参考点(ARP),平均相位中心,相位中心偏差(PCO)和相位中心变化量(PCV)的概念^[10]。天线相位中心示意图如图 2.12 所示。

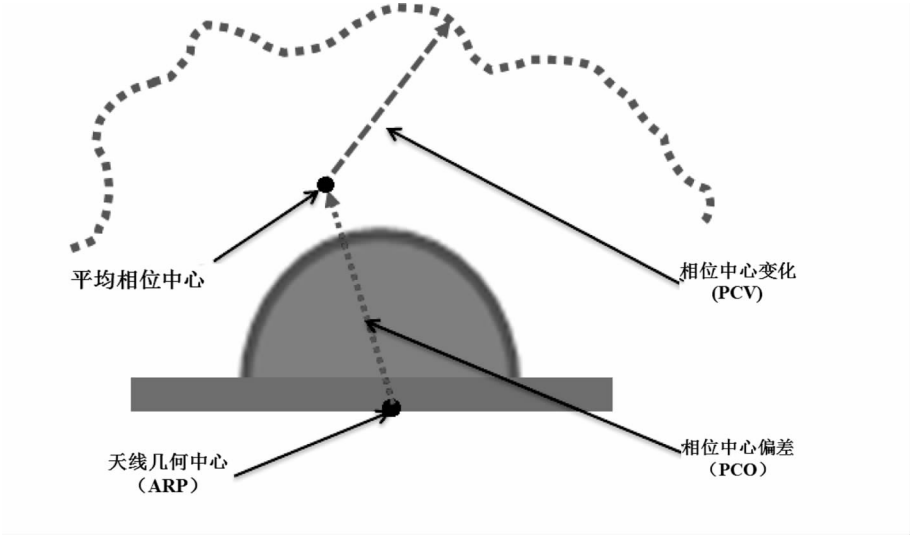


图 2.12 天线相位中心示意图

在空旷的楼顶搭建天线相位中心建模环境,如图 2.13 所示。建立两个水泥基墩分别作为参考站 a 和待测站 b。其中较高的水泥基墩为站 a,较低的为站 b。两站之间距离等于 8 米,构成了超短基线。采用旋转标定法与交换天线法结合的方式标定天线的 PCO^[11-13];在已知天线相位中心偏差 PCO 的基础上,利用室外相对标定法求解天线 PCV^[14-16]。



图 2.12 天线相位中心建模环境

最后将求解出待测天线的 PCO 和 PCV 数学模型修改成 IGS 天线相位中心模型的标准格式,如表 2.1 所示。

表 2.1 待测天线 PCO 标定结果

频率		实测值(mm)
L1	PCO (N)	1.31
	PCO (E)	0.89
	PCO (U)	13.36
	PCV(0-80 度,间隔为 5 度)	0.0, -0.4, -1.6, -2.9, -4.6, -6.3, -7.3, -8.4, -8.8, -9.2, -8.9, -7.9, -6.3, -4.4, -1.6, 1.2, 4.9

频率		实测值(mm)
L2	PCO (N)	1.03
	PCO (E)	-0.32
	PCO (U)	19.82
	PCV(0—80 度,间隔为 5 度)	0.0, -0.1, -0.4, -1.2, -1.9, -2.5, -3.5, -4.1, -5.0, -5.4, -5.8, -5.5, -4.9, -3.7, -2.2, 0.4, 3.2

参 考 文 献

- [1] 李克昭,韩梦泽,孟福军. 北斗系统的特色、机遇与挑战[J]. 导航定位学报, 2014,(5):21—25
- [2] 邢磊. 北斗二代导航天线的仿真与研究[J]. 中国新通信, 2014,(23):111—112
- [3] IEEE Std 145—1993. IEEE Standard Definitions of Terms for Antennas. June 22, 1983
- [4] Luk K M, Wong H, "A new wideband unidirectional antenna element"[J], Int. J. Microw. Opt. Technol. Lett, Vol. 1, No. 1, 2006, pp: 35—44
- [5] An W X, Wong H, Lau K L, et al. Design of broadband dual-band dipole for base station antenna[J]. Antennas and Propagation, IEEE Transactions on, 2012, 60(3): 1592—1595
- [6] Lei Ge, Kwai Man Luk, "A low-profile magneto-electric dipole antenna"[J], IEEE Trans. Antennas Propag, 60 (4), Apr. 2012, pp: 1684—1689
- [7] An W X, Lau K L, Li S F, et al. Wideband E-shaped dipole antenna with staircase shaped feeding strip[J]. Electronics letters, 2010, 46(24): 1583—1584
- [8] Kildal P S. Artificially soft and hard surfaces in electromagnetics[J]. Antennas and Propagation, IEEE Transactions on, 1990, 38(10): 1537—1544
- [9] 范士杰,郭际明,彭秀英. TEQC 在 GPS 数据预处理中的应用与分析[J]. 测绘信息与工程, 2004,(2):33—35
- [10] Wübbena G, Schmitz M, Menge F, et al. A new approach for field calibration of absolute GPS antenna phase center variations[J]. Navigation, 1997, 44(2): 247—255
- [11] 蔡宏翔,王昆杰. 双频 GPS 接收机天线相位中心的测定[J]. 陕西天文台台刊, 1994,(17):25—31
- [12] 郭金运,徐泮林,曲国庆. GPS 接收机天线相位中心偏差的三维检定研究[J]. 武汉大学学报(信息科学版), 2003,(4): 448—451
- [13] 陈逸群,刘大杰. GPS 接收机天线相位中心偏差的一种检定与计算方法[J]. 测绘通报, 2000,(12):15—16
- [14] Schmid R, Rothacher M. Estimation of elevation-dependent satellite antenna phase center variations of GPS satellites[J]. Journal of Geodesy, 2003, 77(7—8): 440—446
- [15] Menge F., Schmitz M. AOADM_T Choke Ring Antenna Absolute Phase Center Variations——Results of Absolute PCV Field Calibrations at IfE and Geo+++. 2001
- [16] Rothacher M, Mader G. Receiver and satellite antenna phase center offsets and variations[C] Data and Analysis Center Workshop. 2002

Design of Antenna for High-Accuracy BDS Application

Shi Luyao

(School of Electronic and Information Engineering, North China Univ. of Tech., Beijing, 100144)

Abstract Beidou satellite navigation system (BDS) is the development of an independent, compatible open global satellite navigation system. With the construction and improvement of the BDS system, the system is moving in the direction of the development of high-accuracy. BDS antenna is an important component, which determines the measurement accuracy of BDS system. Antenna phase center variations and multipath interference are the two main sources of error precision measurement, also a focus and a difficulty of the design of high-accuracy satellite navigation antenna. In this paper, starting with the design of high-accuracy antenna for BDS system, in-depth study of the wideband circularly polarized magneto-electric (ME) dipole antenna, antenna multipath interference suppression techniques and methods to establish the antenna phase center model. The main contents include the following three aspects:

1. The ME dipole antenna, a new complementary antenna element, is studied, and the study is applied to design of the BDS antenna, presenting a novel wideband circularly polarized ME dipole antenna. The proposed antenna consists of crossed ME dipoles, two reflectors and Composite Right/Left-hand (CRLH) power divider with phase shifter. The upper part of the dipole is electric dipole, which is connected to four vertical shorted patches that equivalent to a magnetic dipole. Measured results show that the antenna achieves a wide impedance bandwidth, which cover B1, B2, B3 band of Beidou-2, and in the operating band, the antenna has a good radiation pattern performance and a stable phase center. The antenna has excellent performance makes it completely accords with the design requirements of high- accuracy BDS antenna.

2. Study of the inhibition of multipath interference antenna technology. Focuses on the choke ring technique, multipath interference suppression principle of choke ring was studied. A choke ring is designed for the BDS antenna, and the proposed high-accuracy BDS antenna is mounted on the choke ring. The results show that the improved choke valve after low radiation performance than the elevation axis of the antenna, the antenna is reduced, thereby enhancing the ability of the antenna multipath interference suppression.

3. To study the establishment of the antenna phase center model approach. Reference to foreign high-accuracy GPS antenna phase center modeling method, the phase center model based on international standards of global satellite positioning system service organization (IGS), the establishment of a high-accuracy phase center model.

Key words electromagnetic dipole antenna; circular polarization; multipath; antenna phase center; the beidou satellite navigation system

国际空间站环控生保系统建模与仿真

研究生 王永皎

导师 曲洪权

(北方工业大学电子信息工程学院 北京 100144)

摘 要 国际空间站温湿度控制及通风子系统与航天员的生命安全、生活环境及工作效率紧密相连,是环控生保系统的一个重要子系统。因此,有必要建立一整套国际空间站环控生保系统的精确仿真平台,完成实验数据积累、设计和控制方案优化、调试时间的压缩。

本文在研究系统结构、组件特性、系统稳定性的基础上,对环控生保系统中温湿度控制及通风子系统进行仿真并对仿真结果进行了深入的分析,仿真是基于 MATLAB/Simulink 平台。首先,研究系统工质的物性,推导出这些物性参数的拟合公式,建立系统的物性模型。其次,为每个设备建立压力流量模型。最后,研究克服网络压力与流量耦合导致的系统仿真不稳定问题。

论文着重研究特征值与稳定性、特征值与响应时间之间的关系,研究用微分方程刚性解来解决数值仿真中代数环导致的发散问题。

根据研究结果在 Simulink 中建立简化系统和完整复杂系统的模型,实现了任意变结构条件下的稳定性。

关 键 词 温湿度控制及通风净化系统;物性;动态建模仿真;系统稳定性

作者简介 王永皎,男,1988 年出生,硕士研究生,主要研究方向为系统建模与仿真

E-mail: wyj5550422@126.com

1 绪论

国际空间站的环境控制与生命保障系统是使人在外太空生存的一系列设备的总称,主要分为以下几个分系统:环境控制分系统、大气再生分系统、水的供应和处理分系统、废物处理分系统、热控分系统、居住分系统和舱外活动分系统。在环境控制分系统中每个分系统又可以划分成多个子系统:大气温度控制、湿度控制、压力控制、成分控制、通风、大气监测、复压控制、壁温控制和泄露补偿等分系统。在本课题中我们注重研究环境控制分系统中的通风净化控制通风子系统。

环控生保系统属于复杂的闭环系统,设备间彼此相互影响,系统运行性能影响参数多,设计复杂。温湿度控制及通风系统中,管道中工质的压力和流量之间相互耦合,流量会影响温度,温度又会影响压力和流量的变化,温度同时影响大气中的湿度变化,各个设备之间的温度通过工质也会相互影响。如果还是按照传统的设计方法,测量的数据多,数据量大,工作的难度也很大。因此需要计算机仿真技术,改善原有传统设计的方法的局限性,从动态和系统的角度对各个设备进行最佳匹配设计,并通过计算机仿真试验验证系统设计方案,提高系统设计质量,提高设计效率,缩短设计周期,改善系统运行性能。

2 正文

2.1 建模规则

在仿真平台建立模型时,需要注意的一些原则:

(1) 每个设备的模型建立要考虑到这个系统的运行状态,既要体现该设备元件的自身特性,又能兼顾系统运行的稳定性和可靠性。

(2) 采用模块化建模的方法,将系统分块,分别建立数学模型,然后再将有关模块连接求解出系统模型。

(3) 在流体网络中,压力和流量参数之间存在着水力耦合关系,而压力具有双向传递性,划分模块的时候就解决水力耦合的问题。

(4) 基本模块仿真模型在各设备具体数学模型的基础上,在 MATLAB/Simulink 仿真进行特殊设计。

(5) 在系统中,各种参数的单位要一致,一般用到的参数单位:温度—K,压力—Pa,质量流量—kg/s,时间—s,功—kJ,功率—W,长度—m,焓值—kJ/kg,摩尔质量—g/mol。

2.2 组件分析与建模

课题中分析温湿度控制及通风子系统的结构组成,内部组件结构包含 THCS1 和 THCS2 两部分,THCS1 中主要包含一条主路和一条支路,THCS2 中包含一条主路和两条支路。

根据第二部分中提及的原则对系统中工质物性进行建模,并放入系统模型库中,并对建立的模型库进行管理。建立的系统主要用到的组件模型包括风机、Y 型分三通、Y 型合三通、冷凝换热器、分 8 通、阀门、管道等。分析各个组件的质量守恒方程、动量守恒方程和能量守恒方程。接下来分析要建立系统的稳定性理论,并找到控制系统稳定运行的方法。

2.3 系统稳定性研究

系统的稳定性是系统自身固有特性,与系统的外部输入无关。对不同系统,有不同的理论方法,如果系统是定常线性系统,也就是 $\dot{X}=AX$,那么对于该系统,当 A 为非奇异矩阵时, $X_e=0$ 是系统的唯一平衡状态;当 A 为奇异矩阵时,对于 $AX=0$ 有无数的解,即系统是不稳定的。对于线性定常(LTI)系统,一般认为 X_e 的稳定性代表系统的稳定性。如果系统不是线性系统,那么系统的稳定性需要用到李亚普诺夫(Lyapunov)稳定性理论分析,设系统的其次状态方程为

$$\dot{X}=f(X,t) \quad X(t)|_{t=t_0}=X_0$$

其中, $X(t)$ 为系统的 n 维状态向量, $f(X,t)$ 为 n 为矢量函数,不一定是线性定常的,系统要用到微分方程或者传递函数分析,也就是用经典控制理论来分析,也可以用不求解系统方程的方法,通过用李亚普诺夫函数的标量函数直接判断系统的稳定性。寻找数值算法分析数值算法绝对稳定域,找到控制特征值的数值方法。

第一,转换方程组,得到微分方程组。分析实际系统的特性,得到系统的线性代数方程,对线性代数方程进行建模仿真时,分析代数方程对应的系数矩阵的特性,对于线性代数方程 $Ax=b$,我们分析系数矩阵的特性时发现,其条件数 $cond(A)$ 非常大,而且从得到的系数矩阵看,其中当实际系统中阀门关闭或打开,风机打关闭等操作时,系数矩阵中的某一些元素会发生大的变化,导致矩阵特征值发生剧烈变化,系统呈现病态线性。因此,系统为病态系统,对应的方程组为病态线性方程组,而且在 MATLAB/SIMULINK 平台仿真中,面临着代数环的困扰,为此,将病态线性方程组转为刚性微分方程组的

形式进行仿真。分析刚性微分方程组系数矩阵的特征值,若线性系统中 A 的特征值 λ_j 满足条件 $\operatorname{Re}(\lambda_j) < 0 (j=1, 2, \dots, N)$, 且

$$S = \frac{\max_{1 \leq j \leq N} |\operatorname{Re}(\lambda_j)|}{\min_{1 \leq j \leq N} |\operatorname{Re}(\lambda_j)|} \gg 1$$

刚性比 $S \gg 1$ 时, A 为病态矩阵。通常 $S \geq 10$ 就是刚性的。 S 越大病态越严重。

第二,求得微分方程组系数矩阵。对于常微分方程刚性方程组,进行数值仿真时,有初始值选择问题,也要考虑系统的稳态解问题。对于 N 阶微分方程组

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_n) \quad (i=1, 2, \dots, n)$$

在奇点 x_0 附近对函数 $f_i(x_1, x_2, \dots, x_n)$ 进行泰勒展开

$$\begin{aligned} & f_i(x_1, x_2, \dots, x_n) - f_i(x_1^0, x_2^0, \dots, x_n^0) \\ &= \sum_{j=1}^n \frac{\partial f_i}{\partial x_j} \cdot (x_j - x_j^0) + O(|x - x_0|^2) \end{aligned}$$

我们会得到如下所示的线性函数与奇点处的二阶无穷小之和的方程组,称为微分方程组在奇点 x_0 处的线性化算子(或导算子)。线性部分为 $\dot{x} = Df(x_0) \cdot (x - x_0)$ 我们称线性方程组为非线性方程组的线性化方程。方程组的特征矩阵如下所示,

$$Df(x_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x_0) & \frac{\partial f_1}{\partial x_2}(x_0) & \frac{\partial f_1}{\partial x_n}(x_0) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(x_0) & \frac{\partial f_n}{\partial x_2}(x_0) & \frac{\partial f_n}{\partial x_n}(x_0) \end{bmatrix}$$

泰勒展开就是将非线性系统进行线性近似。涉及到的知识包含一类非线性平面自治系统,几乎线性系统解的稳定性等。对得到的系数矩阵,同样求得系数矩阵的特征值,并分析系数的特征值。

第三,分析微分方程组系数矩阵特征值。要使连续系统稳定运行,需要满足系统稳定的定义。线性稳定和李雅普诺夫稳定性理论。对求得的特征值,用李雅普诺夫稳定判别理论进行分析。李雅普诺夫稳定性判别方法:

定理 1:如果微分方程组在奇点的导算子的所有特征值的实部都是小于零的,则该方程组在点是局部渐进稳定的。

定理 2:若奇点的导算子的特征值中至少有一个特征根的实部为正,则该奇点是不稳定的。

对于数值仿真,也就是离散系统的仿真时,其对应的稳定域会发生很大的变化。相比较而言,稳定域变得特别苛刻。

第四,寻找数值仿真方法和特征值稳定域。通常,只考察数值方法本身时,只检验将数值方法用于解模型方程的稳定性,模型方程为: $y = \lambda y$ 其中 λ 为复数。对一般方程可以用局部线性化转化为这种形式,例如在 (\bar{x}, \bar{y}) 的邻域,可展开为

$$y' = f(x, y) = f(\bar{x}, \bar{y}) + f'_x(\bar{x}, \bar{y})(x - \bar{x}) + f'_y(\bar{x}, \bar{y})(y - \bar{y}) + \dots$$

略去高阶项,再做变换即可得到 $\mu' = \lambda \mu$ 的形式。对 m 个方程的常微分方程组,可线性化为 $y' = Ay$, 这里 A 为 $m \times m$ 的雅可比矩阵 $(\partial f_i / \partial y_i)$, 若 A 有 m 个特征值 $\lambda_1, \lambda_2, \dots, \lambda_m$, 其中 λ_i 可能是复数,所以为了使通过模型方程得到的结果能推广到常微分方程组,方程(4-13)中 λ 为复数,为保证微分方程本身的稳定性,还应假定 $\operatorname{Re}(\lambda) < 0$ 。同时,对于数值仿真,我们分析了微分方程组数值仿真对应的初值问题,以及数值仿真中步长与特征值和绝对稳定域之间的关系,确定数值仿真选择定步长方式。

在论文中分析了不同数值仿真系统的稳定域,分析了欧拉法、一阶到四阶龙格库塔法、隐式欧拉法和梯形法对应的绝对稳定域。如下图所示。

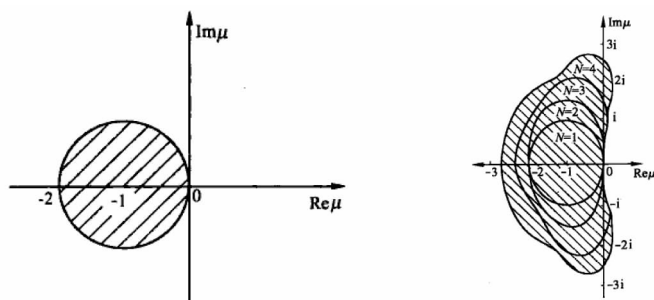


图 1 欧拉法和龙格库塔法绝对稳定域

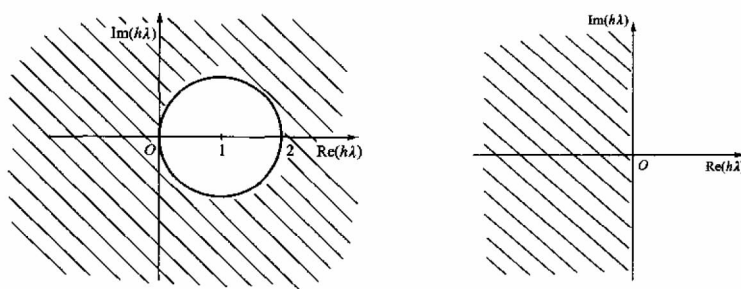


图 2 隐式欧拉法和梯形法绝对稳定域

为了满足系统需要,特征值都要在坐标的左半平面,对于数值仿真,特征值不一定越大越好,结合 Simulink 中给定的数学算法,最终确定选择欧拉法。

第五,找到控制特征值在稳定域内的方法。对于定步长数值仿真系统,当操作实际系统时,系统结构发生变化,对应到数学仿真中系数矩阵发生变化,也就是系统的特征值发生变化,为此我们需要找到控制系统特征值在稳定域内的方法。基于电路理论的自适应虚拟电容方法。

对于该系统中,风机看作是升压器件,对应电路中的电压源,在流体中风机对气流具有阻力,称之为流阻,其流阻对应电路中的电阻;阀门的流阻对应于电路中的电阻,开关阀门相当于增加减小电阻的阻值,导致系统变结构。同理,系统中的管道、冷凝换热器、分三通和合三通等也同样具有流阻,可以看作电路中的电阻器件。那么建立系统对应的电路模型,根据电路理论可以得出对应的电路系统的线性代数方程组,可以求出对应的系数矩阵,当其中一个电阻发生突变时,或者相当于开关突然断开和突然关闭时,系统的系数矩阵中的某一项元素值发生巨大的变化,系统呈现病态。

同时,我们发现对应的电路中没有动态元件,也因此仿真过程中,我们根据之前分析的理论知识,我们将系统中的每一个器件加入相应的虚拟电容,例如,风机用电压源与电阻串联后再与虚拟电容并联的方式表示,阀门用电阻与虚拟电容并联的方式表示。这样我们就可以得到动态系统,再根据电路理论我们可以得到微分方程组,然后求得微分方程组的系数,并确保微分方程组的系数特征值在绝对稳定域内。

因此,我们可以推导出虚拟电容的值跟其并联的电阻值相关,因此在对应的方程组系数中乘以对

应的与电容值相关(或者说电阻值相关的)模型系数,这样确保对应的系统能够稳定运行。

对应的系统系数矩阵的变化情况,由如下所示的矩阵 A 变为矩阵 M。

$$A = \begin{bmatrix} -1 & 0 & 0 & -R_1 \\ 1 & -1 & 0 & -R_2 \\ 0 & 1 & -1 & -R_3 \\ 0 & 0 & 1 & -R_4 \end{bmatrix} \rightarrow M = \begin{bmatrix} -\frac{1}{A} - \frac{1}{B} & \frac{1}{B} & \frac{1}{B} \\ \frac{1}{B} & -\frac{1}{B} - \frac{1}{C} & \frac{1}{C} \\ 0 & \frac{1}{C} & -\frac{1}{C} - \frac{1}{D} \end{bmatrix}$$

2.4 建立系统模型进行仿真

根据理论分析和找到的控制特征值方法最终实现系统的建模与仿真。首先是建立模型库,建立的模型库有统一规定接口。再根据建立的模型搭建对应的一些简单系统,然后再搭建我们的实际系统,并对系统进行校对优化。

首先建立简单的系统模型进行仿真,测试系统的运行效果,经过测试结果为系统稳定运行,结果正确。下图为简单系统物理结构图。

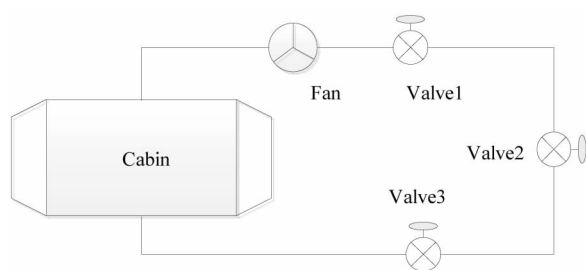


图3 简单系统物理结构图

然后再建立如论文附录4中所示的复杂的温湿度控制及通风系统(THCS),并对其进行仿真,对支路进行操作,然后确定系统是否稳定运行,经过仿真发现,系统能够稳定运行,结果正确。下图为温湿度控制及通风系统结构图。

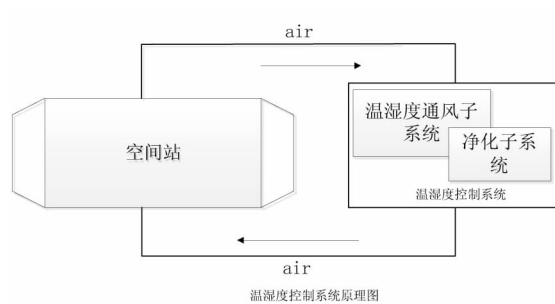


图4 温湿度控制及通风系统结构图

参 考 文 献

- [1] 黄志德,沈学夫.空间站环境控制与生命保障技术[J].中国航天,2000,2:28—32
- [2] 中国人民解放军总装备军事训练教材编辑工作委员会.航天环境控制与生命保障工程基础[M].北京:国防工业出版社,2004
- [3] Wieland, P. O. Living together in space: the design and operation of the life support systems on the International Space Station, NASA TM 98—206956(V01.1) [R]. Marshall Space Flight Center, 1998 • —01: 183—186
- [4] Daues, K. A history of spacecraft environmental control and life support systems[R]. NASA Johnson Space Center, 2006—01—06
- [5] 陈桂明,张明照,戚红雨等.应用 MATLAB 建模与仿真[M].北京:科学出版社,2001
- [6] 薛定宇,陈阳泉.基于 MATLAB/Simulink 的系统仿真技术与应用[M].北京:清华大学出版社,2011
- [7] 陈新宇.基于 HLA 的通信与通信对抗一体化训练仿真系统研究[D].成都:四川大学,2006
- [8] 陈宗海.过程系统建模与仿真[M].合肥:中国科学技术大学出版社,1997
- [9] 董素君.高低温环模系统 MATLAB 仿真模型库搭建技巧[D].北京:北京航空航天大学,2005
- [10] 童钧耕,卢万成.热工基础[M].上海:上海交通大学出版社,2003
- [11] 刘辉,项昌乐,郑慕侨.模块化建模方法及其在车辆传动系中的应用[J].北京:北京理工大学,2002,3,: 61—66
- [12] 刘光启等.化工物性算图手册[M].青岛:化学工业出版社,2002
- [13] 陈则韶,胡芑,程文龙.饱和蒸气密度、焓和蒸发潜热的通用对比态推算式[J].工程热物理学报,2003,24(2): 198—201
- [14] 刘志刚,刘咸定,赵冠春.工质热物理性质计算程序的编制及应用[M].北京:科学出版社,1992
- [15] Rose, S; Fourquet, H; Rolland, A. Columbus orbital facility condensing heat exchanger and filter assembly [C]//29th International Conference on Environmental Systems(ICEC). Denver, Colorado, 1999—07—12
- [16] 王寒栋.泵与风机[M].北京:机械工业出版社,2009
- [17] 蔡增基,龙天渝.流体力学泵与风机[M].北京:中国建筑工业出版社,1999
- [18] 王浚,黄本诚等.环境模拟技术[M],北京:国防工业出版社,1996
- [19] He Jun, Zhao Jingquan. Dynamic simulation of the aircraft environmental control system [J]. Chinese Journal of Aeronautics,2001, (3): 129—133
- [20] 王正林,王胜开等.MATLAB/Simulink 与控制系统仿真[M].北京:电子工业出版社,2012
- [21] 杨洋.关于交流变频异步电力测功机系统的仿真研究[D].长沙:湖南大学,2001
- [22] 沈绍信.线性系统理论[M].大连:大连理工大学出版社,1989
- [23] 钱祥征,戴斌祥,刘开宇.非线性常微分方程理论方法应用[M].长沙:湖南大学出版社,2006
- [24] 宋召青等.基于欧拉算子的劳斯判据[J].烟台:海军航空工程学院,2000,4,626—628
- [25] 钱慧芳.非线性系统的一种线性控制方法研究[D].西安:西安工程大学,2006
- [26] 廖道争,张明廉.一类非线性系统的鲁棒自适应控制[J].北京:北京航空航天大学,2006,18(9),2508—2540
- [27] 刘次华.随机过程[M],第四版.武汉:华中科技大学出版社,2008
- [28] Witt, J; MOiler, R; Bockstahler, K. Columbus ECLSS temperature and humidity control[C]//29th International Conference on Environmental Systems(ICES). Astrium GmbH, Germany, 2010
- [29] Althausen, D. M; Quinn, F. D; Bunnell, C. T. Porous media condensing heat exchanger with integral gas liquid separation for space flight use[C]//44th AIAA Aerospace Sciences Meeting and Exhibit. Reno, Nevada, USA, 2006—01

- [30] Hasan, M. M; Khan, L. I; Nayagam, V. Conceptual design of a condensing heat exchanger for space systems using porous media[C]//35th International Conference on Environmental Systems(ICES). Rome, Italy, 2005—07
- [31] 卜瑶瑶,曹军,杨晓林. 国际空间站舱内空气温湿度控制技术综述[J]. 航天器环境工程,2013, 30(1): 20—25
- [32] 袁兆鼎,费景高,刘德贵. 刚性常微分方程初值问题的数值解法[M]. 北京:高等教育出版社,1992
- [33] 姚征,钟万勰. 椭圆函数的精细积分改进算法[J]. 大连:大连海事大学,2008,4,251—260
- [34] 周义仓等. 常微分方程及其应用[M]. 北京:科学出版社,2003
- [35] 李庆扬. 常微分方程数值解法(刚性问题与边值问题)[M]. 北京:高等教育出版社,1992
- [36] 高普云. 非线性动力学[M]. 长沙:国防科技大学出版社,2005
- [37] 李庆扬,王能超,易大义. 数值分析[M],第五版. 北京:清华大学出版社,2008
- [38] 邱杰,袁渭兰. 数字计算机仿真中消除代数环问题的研究[J]. 计算机仿真,2003,20(7): 33—40
- [39] 马晓虹. Matlab 中的代数环问题及其消除方法[J]. 科技广场,2010,7:159—161
- [40] 李颖. Simulink 动态系统建模与仿真[M],第二版. 西安:西安电子科技大学出版社,2009
- [41] Williams, D. E. International Space Station temperature and humidity control subsystem verification for Node 1 [R]. NASA Johnson Space Center, 2007
- [42] Arthur Batalov. Pressure-Flow Network Modeling[R]. Norway: Department of Engineering Cybernetics Faculty of Information Technology, Mathematics and Electrical Engineering, Norwegian University of Science and Technology (NTNU),2009
- [43] 史昌荣,魏丰. 矩阵分析[M]. 北京:北京理工大学出版社,2010
- [44] Tillner R R, Li J, Yokozeki A, et al. Thermodynamic Properties of Pure and Blended Hydrofluorocarbon (HFC) Refrigerants. Tokyo: The Japan Society of Refrigerating and Air Conditioning Engineers, 1998. 135—810
- [45] Fernando L. Alvarado. Solving power flow problems with a Matlab implementation of the Power System Applications Data Dictionary[J]. Decision Support Systems, 2001(30):243—254
- [46] Lars Johansson. A flight dynamics course based on MATLAB computer assignments[J]. Aircraft Design,2000 (3): 249—259
- [47] 付仕明,徐小平等. 空间站乘员睡眠区二氧化碳聚集现象[J]. 北京航空航天大学学报,2007,33(5):523—526
- [48] HUANG Zhide, TANG Lanxiang. Development of environmental control and life support system of spacecraft and analysis of flight result[J]. Spacecraft Engineering, 2004, 3(1): 37—44
- [49] Batalov, A. Pressure-Flow Network Modeling[R]. Norway: Department of Engineering Cybernetics Faculty of Information Technology, Mathematics and Electrical Engineering, Norwegian University of Science and Technology (NTNU),2011

Modeling and Simulation of the ECLSS in International Space Station

Wang Yongjiao

(Electronic Information Engineering College, North China Univ. of Tech., Beijing, 100144)

Abstract Temperature and Humidity Control and Ventilation are important subsystems of Environmental Control and Life Support System in the International Space, because its will influence the life safety, living environment and work efficiency of astronauts. It is necessary that building a simulated platform of International Space and benefits to accumulate data, optimize design and de-

crease development time.

The simulation of subsystem of Temperature and Humidity Control and Ventilation, system structure, module characteristic, stability and simulate outcome are studied based on the MATLAB/Simulink in this paper. First of all, physical properties are studied and corresponding formulas are gotten. The models which adapt to stability theory can be established. Moreover, models of pressure and mass flow can be established for all devices. Finally, we need to do some research to overcome problems of stability which due to coupling between pressure and mass flow.

The relationship between eigenvalue and stability or eigenvalue and response time is the most important content of this paper. The way of solution for stiff ordinary differential equation is proposed to solve problem of diverge, because the system has algebraic loop.

According to the research, simplified and whole system can be established in Simulink. In addition, the stability can be guaranteed in different condition of variable structure.

Key words temperature and humidity control and ventilation system; physical properties; dynamic modeling and simulation; system stability

高密度客流实时检测系统设计

研究生 乔 帅

导师 张常年

(北方工业大学电子信息工程学院 北京 100144)

摘 要 目前,客流检测已成为计算机视觉和智能交通等相关领域研究的热点,其在实际生活中的应用也越来越广泛。在公共安全领域,管理者通过客流监控可以建立客流预警机制,对超出现场所承载力的客流进行有效引导和疏散,从而减少拥挤踩踏事件发生;在商业领域,经营者通过客流量的变化可以适时调整商场经营策略,提升购物环境的舒适性和人力资源调配的合理性;在交通领域,线路规划者根据各站点客流变化规律可以制定出合理的交通线路从而提高交通运输效率。

目前的客流检测技术在目标检测时容易受到光线变化、复杂背景环境等的干扰。特别是在人群密集的环境下,复杂多变的背景环境以及行人之间相互遮挡的情况都会给目标检测带来极大的困难和挑战,现有的客流检测系统在此情况下很难满足实际应用对系统检测精度和实时性的要求。基于此本文设计实现了能够适用于高密度客流环境下的客流检测系统。

在研究了现有行人检测算法的基础上,本文首先采用深度摄像机获取场景的深度图像和彩色图像,其次根据背景差分和基于距离信息的阈值分割方法确定行人头肩部出现的区域,然后对头肩部区域提取梯度直方图特征并采用支持向量机对目标进行分类和识别。在得到目标检测结果后采用 Kalman 与 MeanShift 相结合的方法对行人进行跟踪并获得目标的运动轨迹,然后根据目标的运动轨迹实现双向客流量统计和客流速度、客流密度的估算。此外,除实现客流数据计算功能外,本客流检测系统还实现了视频和数据保存功能以及远程监控的功能。最后通过在地铁站多个客流密集区域的现场测试,验证了本论文设计实现的客流检测系统的稳定性和有效性。

关 键 词 客流检测;深度图;行人检测;HOG;目标跟踪

作者简介 乔帅,男,1989 年出生,硕士研究生,主要研究方向为视频图像处理与智能交通。

E-mail: qshuai89@163.com

1 绪论

1.1 研究背景与意义

近些年随着我国城市化进程的不断加快,城市密集人群的流动越来越频繁,而且规模也越来越大。对于大规模的人群流动,如果不采取有效的监管和安全防范措施,势必会影响社会治安并给公民人身安全带来巨大损失。如何有效的对人群分布进行合理规划,保障个体的人身安全已成为管理者亟待解决的问题。

然而,当面临这样的大规模人群流动时,如果缺乏相关的客流数据信息,相关监管部门就不能对局势进行全面掌控从而错失处置应急事件的最佳时机。因此,提供满足高密度客流状态下客流信息的实时准确获取手段,加强对突发大客流的检测技术研究不仅能提升相关部门快速疏散旅客能力以及迅速制定应急处理方案的能力,而且对我国大型公共管理水平的提升以及对大型公共场所资源的合理分配也具有重大的现实意义。

1.2 国内外研究现状

目前,客流检测系统的研发已成为许多科研工作者研究的热点之一。客流检测技术通常包括以下几种方式:压力传感检测方式、红外传感检测方式和基于视频分析的检测方式等。基于压力传感和红外传感的检测方式不能有效的区分人或物体,同时多人进出时检测的精度会降低,而基于视频分析的方法能克服上述不足,因此得到越来越多的科研工作者的关注。

基于视频分析的客流检测是在行人检测的基础上,通过连续跟踪及轨迹分析进行客流数据计算的过程。国内外科研工作者对此做了大量的工作,特别是在行人检测和目标跟踪两个重要方面。在行人检测方面,基于深度图的行人检测方法充分利用场景的深度信息,从立体的角度来描述真实场景。这种检测方法能够有效减少因光线变化而对检测效果的影响,因而得到众多科学家的关注和研究。在行人跟踪方面,由于行人姿态多样,运动方向不明确,相对于刚体目标而言,行人特征的提取更具挑战性,为此国内外研究学者在行人目标跟踪方面进行了很多探索和研究^{[1][2]}。常用的跟踪算法主要有基于区域,基于特征,基于模型和基于轮廓的跟踪算法。

2 客流检测关键技术研究

基于视频分析的客流检测方法主要包括行人检测和行人跟踪两部分。本章节主要对客流检测系统设计中涉及的与行人检测和目标跟踪相关的算法进行对比分析,从而为后续设计适用于高密度客流环境下的客流检测方法提供理论基础。

2.1 行人检测算法

在行人目标检测之前对图像进行前景目标分割是非常有必要的。前景目标分割的主要目的是从静态图像或视频图像序列中将感兴趣区域从图像中分割出来,使以后的工作更关注于分割出的部分。目前常用的前景目标分割方法有基于阈值分割的方法和基于边缘检测的方法。

在实现前景目标分割之后,需要对分割出的目标进行特征表示,然后通过分类器对提取到的特征进行分类判别。目前常用的行人描述特征主要有边缘特征、形状特征、颜色特征、统计特征或者变换特征等,其中比较有代表性的特征包括 Haar-like 特征、SIFT 特征、Edgelet 特征、HOG 特征等。Haar-like 特征是 Papageorgiou 等人^[3]首先提出的,Viola 等人^[4]在此基础上首次引入积分图概念,提升了图像 Haar-like 特征值计算的效率。Haar-like 特征由白色和黑色矩形模板组成,其特征值是将白色区域累加和与黑色区域累加和相减而得到。SIFT 特征最早是由 Lowe^[5]在 2004 年首次提出的。SIFT 特征由于具有良好的尺度不变和旋转不变性,在复杂背景下对目标检测具有很好的效果,但算法的不足在于特征的计算量较大。Edgelet 特征是 2005 年由 Wu 等人^[6]提出的一种描述轮廓方向的特征。每个 Edgelet 一般是由很少的像素点组成的一条小边,用于描述某个物体的局部轮廓,该特征对光照变化不是很敏感。

常用检测分类器主要有 K-近邻分类器,AdaBoost 分类器和 SVM 分类器等。K-近邻分类器的主要思想是如果一个样本在特征空间中的 k 个最相似的样本中的大多数属于某一个类别,则该样本也属于这个类别。AdaBoost 分类方法是由 Freund 和 Schipare^[7]首次提出的。其核心思想是针对同一个训练集训练不同的分类器,即弱分类器,然后把这些弱分类器集合起来,构造一个更强的最终分类器。

2.2 运动目标跟踪

运动目标跟踪作为计算机视觉领域的一个重要方面,已广泛应用于航空航天、自动控制 and 国防建设等各项领域。目前常用的跟踪算法有基于特征的跟踪算法,基于 3D 模型的跟踪算法,基于变形模

板的跟踪算法和基于区域的跟踪算法^{[8][9]}。常用的目标跟踪算法有:

(1)Kalman 滤波跟踪:Kalman 滤波器是 1960 年由 Rudolf Emil Kalman^[10]提出的一种能够实现对运动系统状态的最小方差误差估计的线性递归滤波器。它首先通过噪声和信号的状态空间模型误差进行建模,将目标跟踪问题转化为求解系统状态后验概率密度估计的问题从而实现对系统状态进行准确估计。

(2)粒子滤波跟踪^[11]:该算法运用的主要思想是随机采样和重要性采样原则,在不明确目标处于哪个区域时,在整个区域随机的分布粒子,然后根据特征相似度测量粒子的重要性,随后在下一次分布粒子时在粒子重要性高的地方分布更多的粒子,在重要性差的区域分布较少粒子以实现对关键区域的重要性重采样。

(3)MeanShift 跟踪:该算法是 Fukunaga 等人^[12]提出的一种非参数概率密度梯度估计算法,并由 Cheng^[13]等将其推广到计算机视觉领域。它不需要任何先验知识,仅根据特征空间中样本数据计算其密度函数值。该算法的流程主要包括目标模型与候选模型建立、相似性度量和目标定位等几部分。

3 客流检测算法设计

本文所设计的客流检测系统是通过视频分析的方法对行人目标进行检测跟踪,然后依据检测跟踪的结果计算监控区域内的双向客流量、客流速度和客流密度等数据。为解决人群拥挤环境下目标之间相互遮挡的问题,本客流检测系统把不易受遮挡的人体头肩部作为研究对象,采用基于深度图像的阈值分割方法,提取人体头肩部区域的 HOG 特征并结合 SVM 分类器进行行人目标检测识别,然后采用 Kalman 与 MeanShift 相结合的跟踪算法对检测到的目标进行跟踪,最后根据目标跟踪结果进行客流数据计算。系统整体算法流程图如图 3.1 所示。

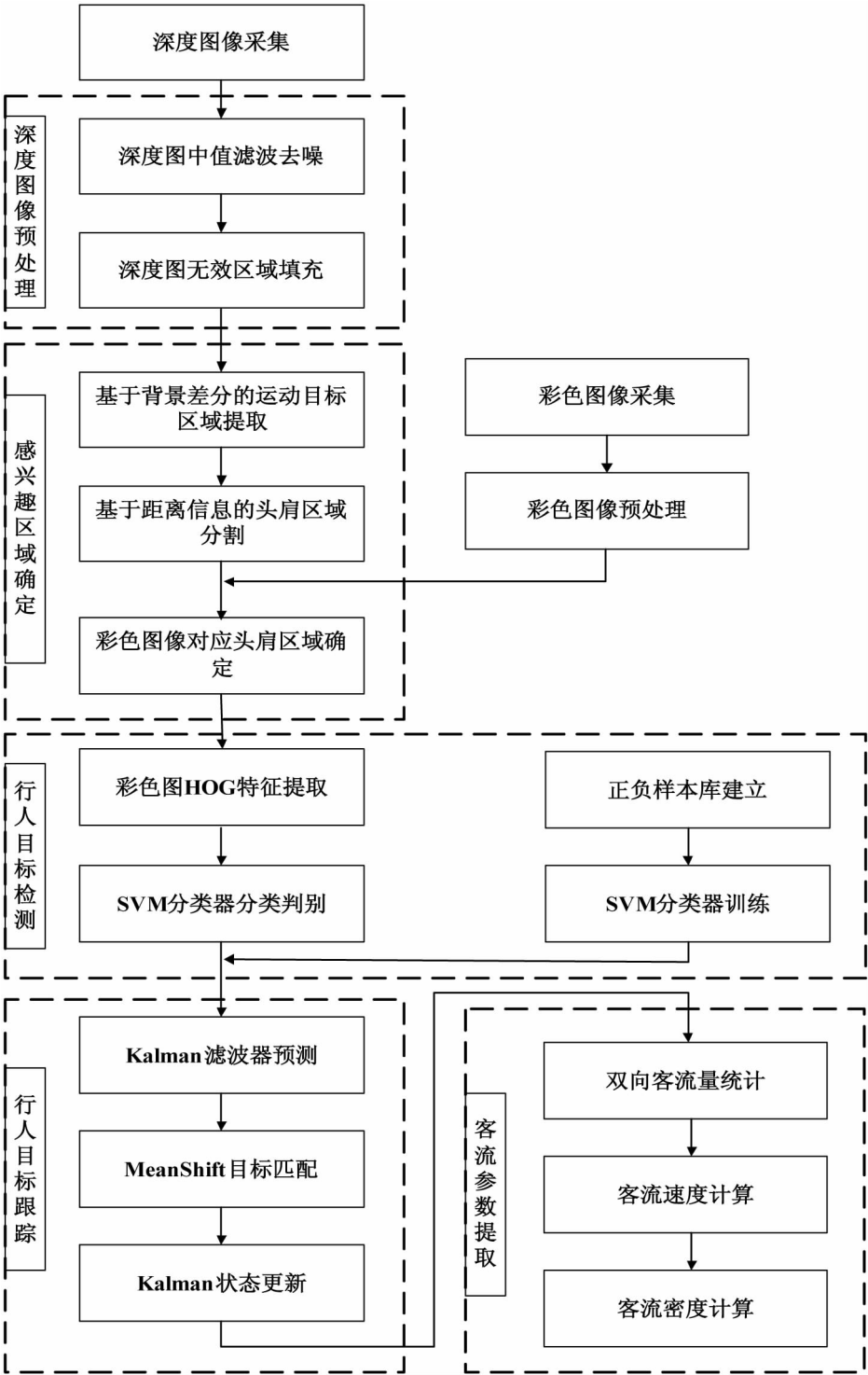


图 3.1 系统整体算法流程图

3.1 图像采集和预处理

本客流检测系统采用一种经过改进的深度摄像机获取监控场景的深度图像和彩色图像。摄像机采用一种被称为光编码技术的结构光方法实现深度图像的采集。由于深度摄像机获取深度图像原理的独特性,在深度图像的某些区域特别是在表面比较光滑的地方和物体的边缘轮廓处会出现黑色无效的区域。为消除这些区域对后面检测跟踪效果的影响,需要对该区域进行适当的处理。为避免这些无

效区域对检测结果的干扰,本文采用最近邻非零值来代替无效区域内各点的值。



(a)原始深度图像



(b)中值滤波后的深度图像

图 3.2 中值滤波前后对比图



(a)无效区域填充前的图像



(b)无效区域填充后的图像

图 3.3 无效区域填充前后对比图

3.2 前景目标分割

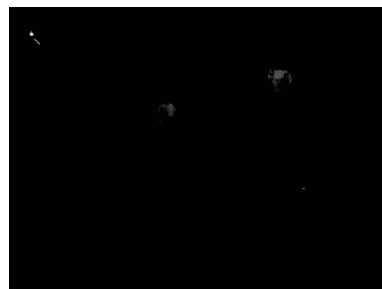
在经过预处理之后,需要对图像进行前景目标分割,其主要目的是从图像中确定可能包含行人的窗口区域,排除场景中其他物体对检测结果的干扰。本客流检测系统采用背景差分法进行前景分割。



(a)均值背景图



(b)含有检测目标的图像



(c)背景差分后的图像

图 3.4 背景差分效果图

通过背景差分能够得到场景中运动物体的整体形态,但是在人群密集的环境中,运动物体之间的遮挡情况相当严重,无法准确区分每个单独的运动物体,因而采用目标整体形态的检测往往不能取得良好的效果。然而通过视频分析发现,行人的头肩部区域在人群拥挤的环境下遮挡现象不是很明显,为此有必要对行人的头肩部区域采取进一步的分割处理。

在本文中主要依据距离信息实现头肩区域分割。假设摄像机能够获取的最远有效距离为 D_{\max} ,灰度图像像素的最大值为固定值 P_{\max} ,根据公式(3.1)可以完成实际距离值与灰度值之间的转换,

$$P = D \cdot P_{\max} / D_{\max} \quad (3.1)$$

其中 D 为各点的深度距离值, P 为对应点灰度值。在完成转换之后,我们就可以通过采用设定灰度阈值范围的方法对行人头肩部区域进行分割,即

$$dst(x, y) = \begin{cases} src(x, y) & MinValue \leq src(x, y) \leq MaxValue \\ 0 & totherwise \end{cases} \quad (3.2)$$

其中 $MaxValue$ 和 $MinValue$ 为根据实际场景设定的最大和最小分割阈值。

3.3 行人目标检测

通过前景目标区域分割,行人头肩部所在的区域已经被提取出来,为进一步识别该区域内的物体是否为行人,需要对该区域进行行人头肩目标检测。在本文中采用 HOG(Histograms of Oriented Gradients)特征与 SVM(Support Vector Machine)相结合的行人检测方法。HOG 特征通过计算和统计图像局部区域的梯度方向直方图作为该局部图像区域的表征,能够对图像的边缘信息、颜色信息等进行综合描述。它将整个图像划分成多个被称为 cell 的较小连接区域,然后计算每个 cell 生成的梯度直方图,最后将这些直方图通过一定方式组合起来构成 HOG 特征描述。HOG 特征计算框图如图 3.5 所示。

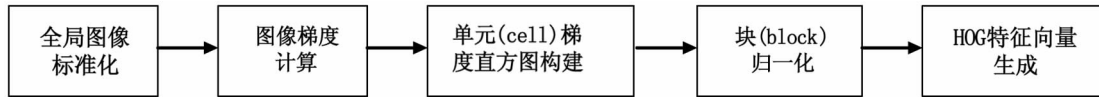


图 3.5 HOG 特征计算框图

在经过 HOG 特征提取后,需要对提取到的特征进行分类判别。SVM 基于结构风险最小化的原则,依据有限的样本信息在训练样本的学习精度与无差错识别任意样本之间寻找最佳的平衡点,以达到最大的泛化能力。SVM 分类器在分类之前,需要通过对一系列正样本和负样本的观察学习,使其对目标具有分类识别的能力。SVM 分类器训练的流程框图如图 3.6 所示。



图 3.6 SVM 分类器训练的流程框图

在经过 HOG 特征提取与 SVM 分类后行人头肩部目标已经得到检测,为了进行客流量统计和客流密度、速度的计算,需要对识别到的行人目标进行跟踪以获得行人完整的运动轨迹。本文采用 Kalman 与 MeanShift 相结合的方法实现对目标的跟踪,在一定程度上解决了目标与背景特征相近以及目

标部分遮挡等对跟踪效果的影响。整个跟踪算法的流程框图如图 3.7 所示。

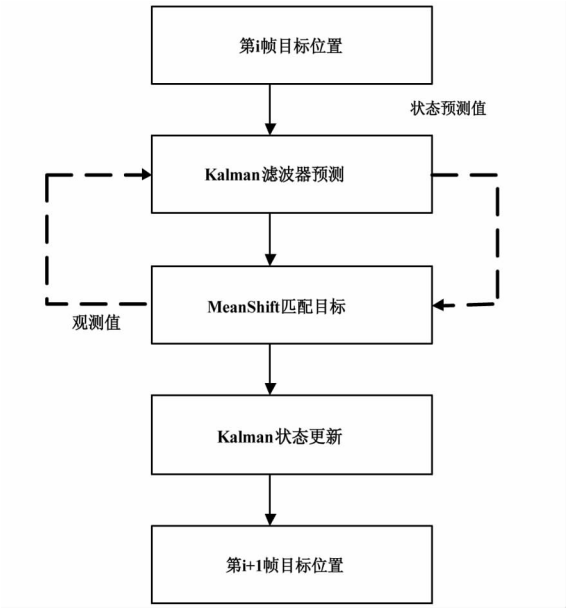


图 3.7 Kalman 与 MeanShift 结合的跟踪算法框图

在得到目标运动轨迹之后可实现对客流量、客流速度和客流密度的计算。客流量的计算通过设置计数线的方法来实现。当行人目标中心点在前一帧中的位置在计数线一侧而目标在当前帧中的位置位于计数线的另一侧时,系统就计数一次。客流速度通过跟踪目标的行走轨迹以及目标在视频序列中出现和消失的时间点来计算。客流密度是通过监控区域内所有目标所占面积在待检测区域面积中的比重来计算。

4 客流检测系统设计与实现

为实现一个能够用于人群密集环境下的客流检测系统,本文以客流密集的地铁换乘站为研究场景,利用空间距离信息和彩色图像信息进行目标检测和跟踪,实现对客流量、客流速度和客流密度等参数的计算,从而为后续客流预警信息的发布和客流监管措施的制定提供数据基础。

系统的整体结构主要由深度摄像机、视频分析专用主机、交换机和服务器组成。深度摄像机采集到图像数据后传输到视频分析主机,视频分析主机根据采集的图像进行视频分析并将采集的监控视频实时存储。视频分析主机通过交换机与服务器相连,服务器对接收到的客流数据进行实时更新和存储。整个系统设备的架构图如图 4.1 所示。

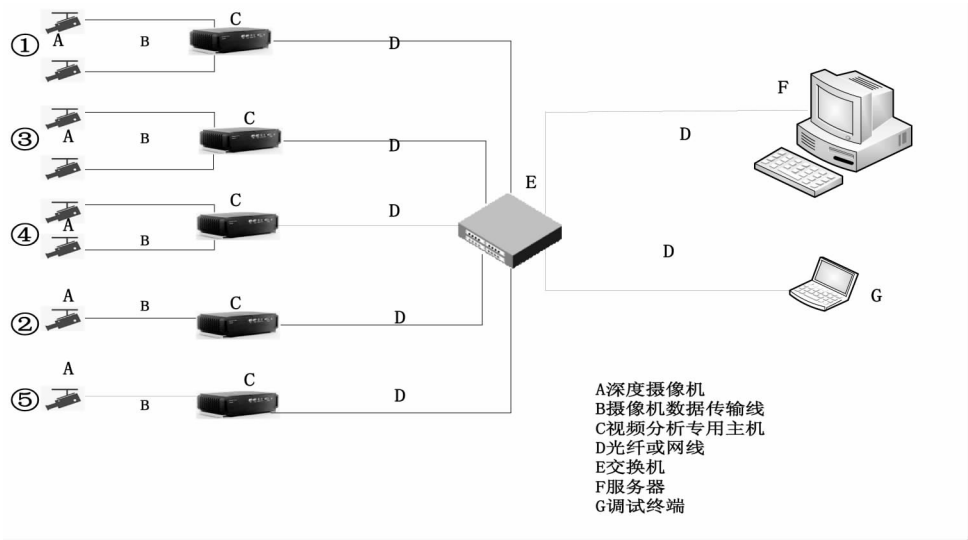


图 4.1 系统设备的整体架构图

客流检测系统可分为客户端和服务端两个部分，发送端部分负责视频图像的获取和处理，目标的检测跟踪，客流量统计，速度、密度估算以及视频信息存储。接收端部分负责远程视频监控画面的切换显示，客流数据信息的接收以及客流数据的存储。系统整体功能模块框图如图 4.2 所示。

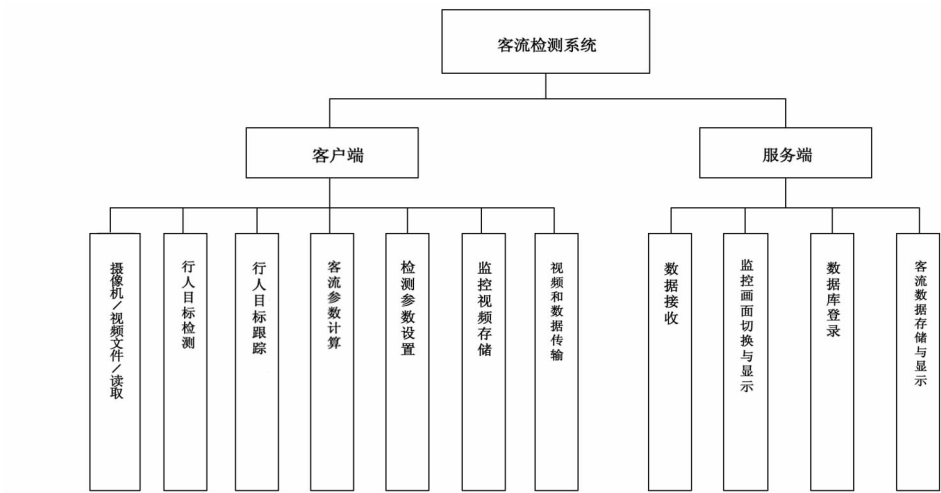


图 4.2 系统整体功能模块框图

为测试本客流检测系统在不同高密度客流场景下的运行效果，本文选取地铁站厅层多个客流较为密集通道口作为测试环境，并对客流高峰时期的多个监控点进行了现场测试，图 4.3、图 4.4 为不同监控点在高密度客流环境下视频截图。

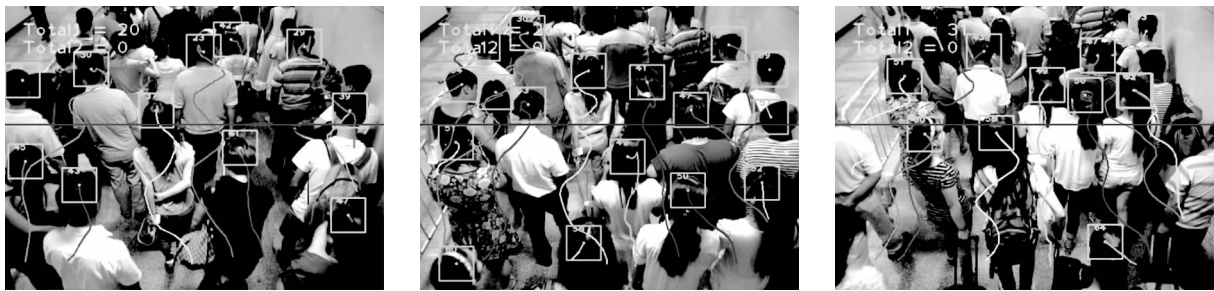


第 971 帧图像序列

第 997 帧图像序列

第 1025 帧图像序列

图 4.3 双向扶梯口处视频截图



第 450 帧图像序列

第 475 帧图像序列

第 500 帧图像序列

图 4.4 换乘通道口处视频截图

表 4.1 为各个监控点在客流高峰时段内的客流量统计表,从统计结果中可以看出,系统在地铁客流高峰时段的客流量统计准确率均在 95.42%以上,可以满足普通地铁站局部监控区域的实时客流统计需求,所提供的实时客流信息可用于建立客流预警平台并作为制定客流交通管理与控制方案的基础和依据。

表 4.1 客流量结果统计表

检测点	视频时长/分	实际人数/人	系统统计人数/人	误检人数/人	漏检人数/人	准确率
一号检测点	5	712	707	11	16	96.21%
二号检测点	7	603	595	7	15	96.35%
三号检测点	4	852	837	12	27	95.42%
四号检测点	8	1091	1098	23	16	96.43%
五号检测点	7	579	572	8	15	96.03%
整体	31	3837	3809	61	89	96.09%

5 全文总结

本文针对客流检测系统在现实生活中的实际需要,在现有客流检测技术研究的基础上,设计并实现了一个基于深度图像和彩色图像信息,能够适用在高密度客流环境下的客流检测系统。该方法首先

根据距离信息提取行人头肩区域,然后对头肩区域提取梯度直方图特征并采用支持向量机对目标进行分类和识别,并采用 Kalman 与 MeanShift 相结合的方法对行人进行跟踪并获得目标的运动轨迹,在得到目标的运动轨迹之后实现双向客流量统计和客流速度、客流密度的估算。本文最后通过地铁站现场环境的实际测试,验证了本客流检测系统的有效性、稳定性和实时性。

参 考 文 献

- [1] 齐飞,罗予频,胡东成. 基于均值漂移的视觉目标跟踪方法综述[J]. 计算机工程,2007,33(21):24—27
- [2] Vidal A, Marron J. J, Labrador M. A. Real-time Pedestrian Tracking in Indoor Environments[C]. Communications (LATINCOM), 2014 IEEE Latin-America Conference on, 2014; 1—6
- [3] Dalal Navneet, Triggs Bill. Histograms of oriented gradients for human detection[C]. Computer Vision and Pattern Recognition, 2005. IEEE Computer Society Conference on. IEEE, 2005, 1: 886—893
- [4] Viola P, Jones M. J, Snow D. Detecting pedestrians using patterns of motion and appearance[J]. International Journal of Computer Vision, 2005, 63 (2): 153—161
- [5] Lowe D M. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91—110
- [6] Wu Bo, Nevatia Ram. Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors[J]. International Journal of Computer Vision, 2007, 75(2): 247—266
- [7] Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of basic Engineering, 1960, 82 (1): 35—45
- [8] 杨戈,刘宏. 实际跟总算法综述[J]. 智能系统学报, 2010, 5(2): 95—105
- [9] 李琦,邵春福,岳昊. 视频序列中面向行人的多目标跟踪算法[J]. 北京理工大学学报, 2013, 22(2): 178—184
- [10] Kalman R E. A new approach to linear filtering and prediction problems[J]. Journal of basic Engineering, 1960, 82 (1): 35—45
- [11] Sabour, N. A. Faheem, H. M. Khalifa, M. E. Multi-agent based framework for target tracking using a real time vision system[C]. International Conference on Computer Engineering & Systems, ICCES 2008; 355—363
- [12] Fukunaga K, Hostetler L. The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition [J]. IEEE Trans. On Inform. Theory, 1975, 21(1): 32
- [13] Cheng . Y . Mean shift, mode seeking and clustering[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 1995, 17(8): 790—799

Design of Real-Time High-Density Passenger Flow Monitoring System

QiaoShuai

(School of Electronic and Information Engineering, North China Univ. of Tech. , Beijing, 100144)

Abstract In recent years, research on passenger flow monitoring system has been being a very active field of computer vision and intelligent transportation. The monitoring system has been applied more and more widely. In public security through monitoring passenger flow can establish warning

mechanism so that reducing accident occurred of swarm and jostlement. In commercial field commercial investor can timely adjust management strategy according to the change of passenger volume so that improving shopping environment and the rationality of the human resource allocation. In the field of transportation traffic route planners can formulate reasonable traffic lines through passenger flow data so that improving efficiency of transportation.

Current passenger flow technology is often affected by many factors such as illumination changes and complex background scene. Especially under the crowded environment, mutual occlusion of moving targets will bring the great difficulty and challenge. In this case, the traditional passenger flow monitoring system is difficult to satisfy the practical requirements of accuracy and real-time. Based on this, this article designs and implements a system which can be applied to crowded environment.

On the basis of researching some pedestrian detection algorithms, this article adopts the method of depth image and color image combination to detect moving pedestrians. It firstly uses the algorithms of combining HOG and SVM to identify targets. Then it adopts MeanShift combined with Kalman filter algorithm to tracking object so that obtaining tracking trajectory. After that, it counts pedestrians and computes velocity and density of passenger flow according to tracking trajectory. In addition to the function of passenger flow data computing, the system also realizes the functions of data conservation and remote monitoring. Finally, by the field test in a subway station, the stability, and accuracy of this monitoring system has been verified.

Key words passenger flow monitoring; depth image; pedestrian detection; HOG; object tracking

光子晶体慢光特性的研究

研究生 权 炜

导师 董小伟

(北方工业大学电子信息工程学院 北京 100144)

摘 要 本文主要针对二维光子晶体能带分布特性、慢光特性和传输特性进行了分析研究。基于数值计算方法,研究不同结构二维光子晶体能带的带隙分布特征;通过对结构的能带色散、群速度分布及群速度色散等波导参数的计算,分析光子晶体缺陷波导的慢光特性以及缺陷介质柱对慢光性能的影响;然后研究频率落在禁带内外的光波在二维光子晶体中的传输特性,计算不同光子晶体缺陷结构的透射谱;最后对耦合腔光子晶体波导结构的慢光特性进行分析,设计了具有高带宽和低群速度的新型耦合腔慢光波导结构,并验证结构具有优良的传输特性。

关 键 词 光子晶体; 滤波器; 慢光; 时域有限差分算法; 平面波展开法

作者简介 权炜,男,1989 年出生,硕士研究生,主要研究方向为光通信器件。

E-mail: qkosmosw@163.com

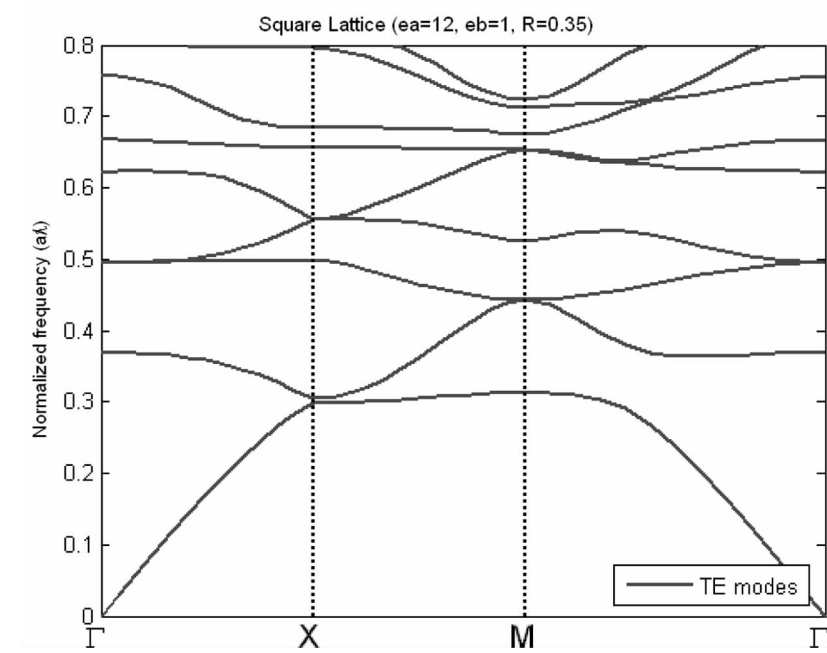
1 绪论

近年来,为了彻底改变光通信系统应用逐渐滞后的窘迫局面,使信息处理及交换路由可以由全光网络实现,全光网络相关技术的研究吸引了越来越多通信领域的研究者。而全光缓存这一科技对网络中的信号调节与交换起着不可或缺的作用。光缓存器是指用来实现光信号传输时的延迟以及调整延迟时间的器件。利用原子气或一些特殊材料可实现光缓存,但造价极高且技术难度很大,无法应用在全光网络中。迄今为止,可以大规模投入使用的全光缓存器还未研制出。光子晶体这一构想的提出,向全光缓存器的研制指明了新的方向。

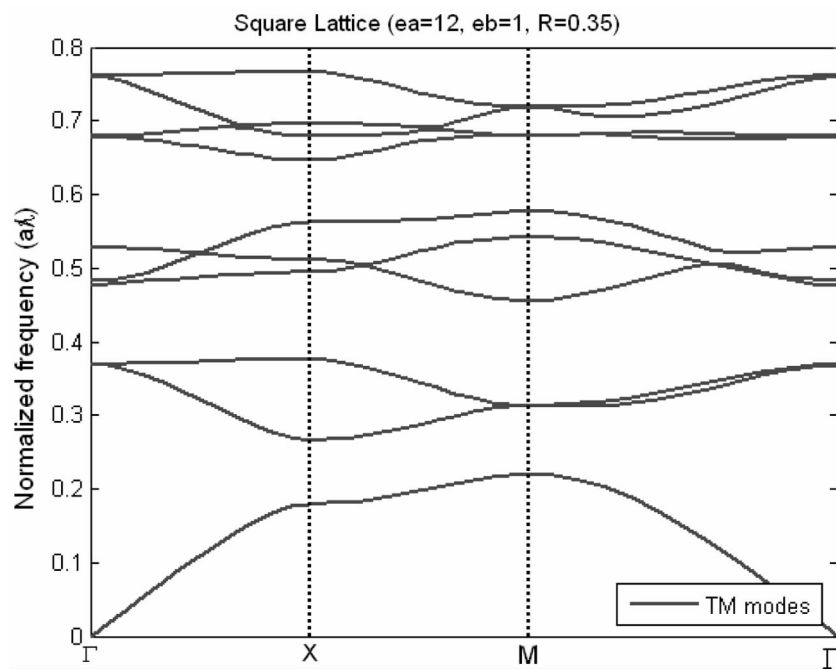
2 二维光子晶体慢光能带特性分析

2.1 二维光子晶体能带分析

正方晶格的光子晶体 TE 模式、TM 模式的能带结构如图 2-1 所示。图中的横坐标表示第一布里渊区不可约波矢分布,纵坐标表示归一化频率,带隙用红色区域标出。



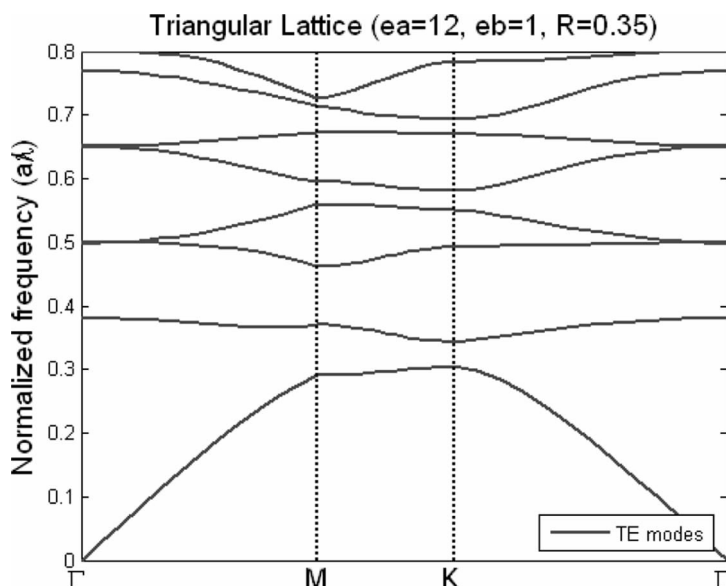
(a) TE 模式



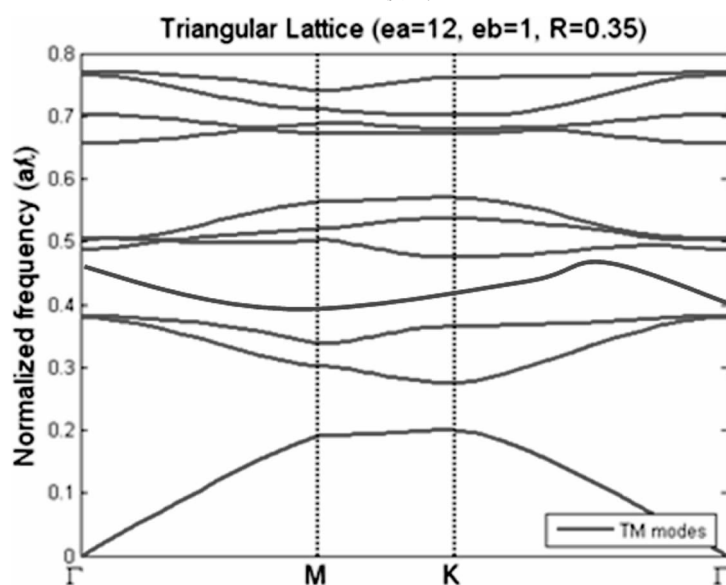
(b) TM 模式

图 2-1 正方晶格光子晶体能带

观察图 2-1(a)可得, TE 模出现一个带隙, 在频率 $0.68-0.69$ 之间, 但带隙宽度很窄。TM 模式存在两处带隙分别在在频率 $0.22-0.27$ 以及 $0.38-0.45$ 之间。TE 模式带隙的频段在 TM 模式第一与第二带隙频段之后, 频率值较低且较窄长, 未发现 TE 模带隙和 TM 模带隙相互重叠的频率区域。



(a) TE 模式



(b) TM 模式

图 2-2 三角方晶格光子晶体能带

根据图 2-2 可得出,红色区域为带隙,三角晶格结构在 TE 模式下存在两处带隙,分别在频段 0.29—0.32 以及 0.40—0.45, TM 模式存在两处带隙分别在频段 0.20—0.27 与 0.58—0.64, TM 模式下得到的两处带隙宽度略大于 TE 模式,TE 模式下的两带隙频段的取值位于 TM 模式中两处带隙频段的取值之间。与正方晶格类似,三角晶格 TM 模式与 TE 模式的能带带隙分布没有发现相互交叠的区域。

2.2 二维缺陷态光子晶体慢光特性分析

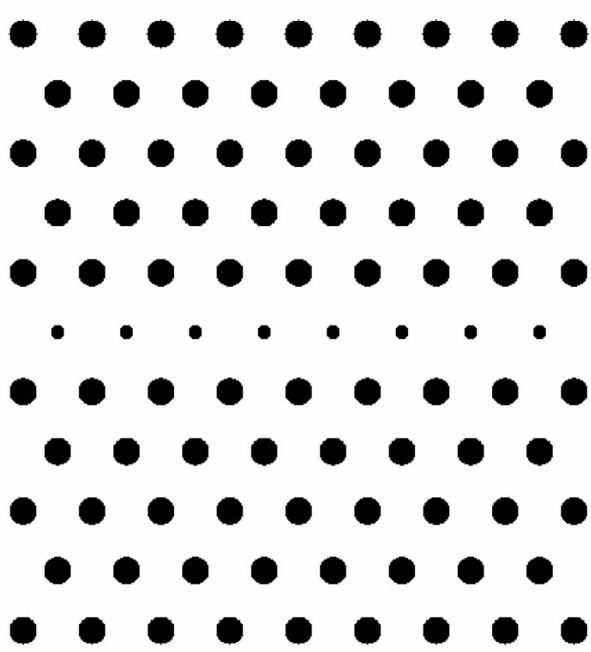


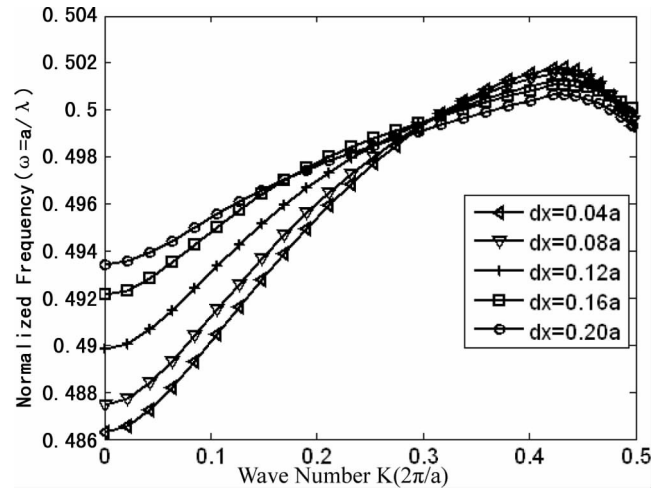
图 2-3 线缺陷三角晶格光子晶体结构

我们以二维三角晶格线缺陷波导结构为例,分析与慢光相关的各项参数,其结构如图 2-3 所示,通过在完整结构中改变某一行介质柱的半径而形成。介质柱材料取值 $\epsilon_a = 8.9$,背景介质为空气 $\epsilon_b = 1$ 。介质柱半径为 $R = 0.2a$,缺陷行介质柱半径为 $R_d = 0.1a$, a 为晶格常数。光波在进入光子晶体后将沿着缺陷行方向传播。

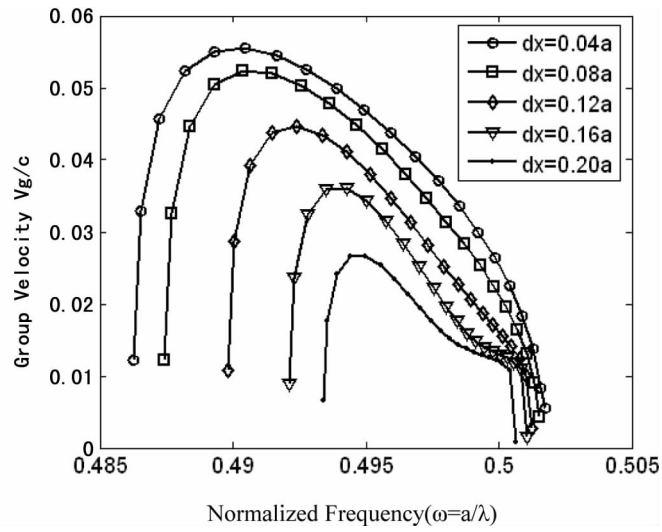
(1) 左右平移

首先固定 y 方向平移量为 0 不变,沿 x 方向左右平移缺陷介质柱,如图 2-4(a)所示。可以看出随着偏移量的增加,导模曲线逐渐向高频范围靠近,同时导模的频率有效带宽在逐渐变窄。而随着偏离量的增大,导模曲线的形状逐渐趋于平坦,出现数值变化缓慢的平稳区域,这说明该区域群速度 v_g 逐渐降低。图 2-4(b)是利用各个导模曲线计算得到的群速度曲线,由图可得,随着偏移量的增大,可以看出随着偏移量的增加,群速度在逐渐下降,但同时慢光的有效频率带宽也越来越窄。

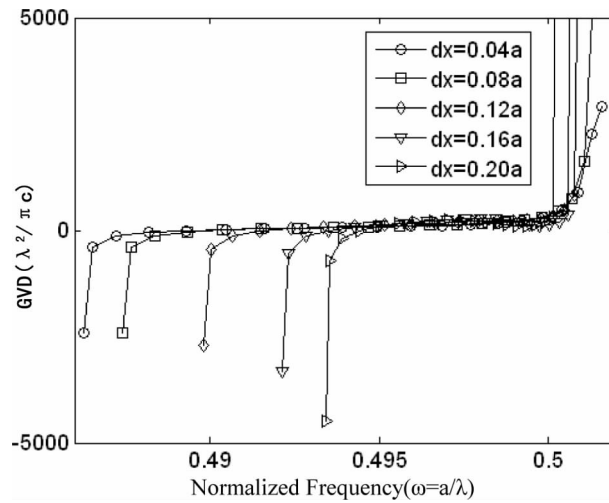
另外还给出了偏移量的变化对群速度色散(GVD)曲线的影响,如图 2-4(c)所示。随着横轴数值的不断增大,GVD 的取值在 y 轴负半轴急剧递增,增大到 x 轴附近时逐渐平坦,之后在进入 y 轴正半轴后曲线又开始急剧递增。可以将导模曲线的尾部区域用作慢光带,该区域在能带色散范围内,群速度取值小,易于慢光器件实现。



(a) 能带色散曲线



(b) 群速度曲线

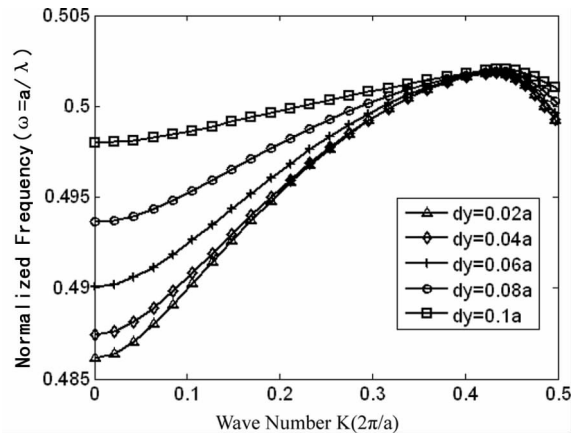


(c) GVD 曲线

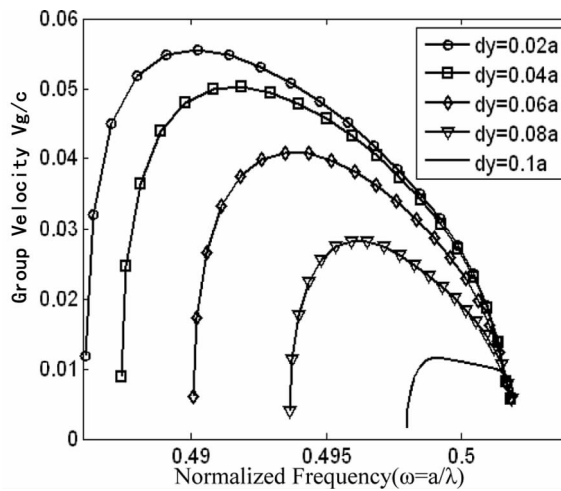
图 2-4 线缺陷三角晶格光子晶体慢光特性随 dx 的变化

(2) 上下平移

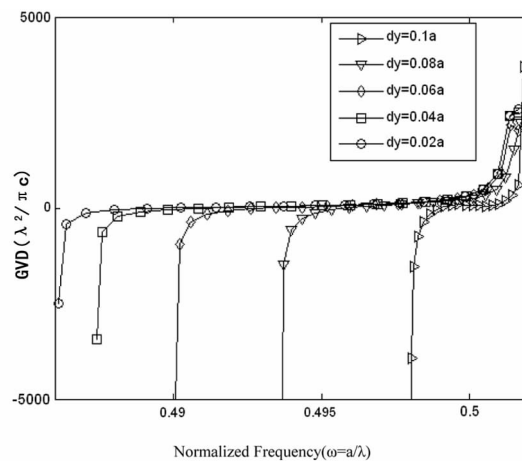
将缺陷行介质柱沿 y 方向上下平移并研究其对导模特性的影响,如图 2-5 所示。导模色散曲线发生上移,结构的群速度随着偏移量增大而逐渐减小。当偏移量 dy 取 $0.10a$ 时,导模色散曲线已接近



(a) 能带色散曲线



(b) 群速度曲线



(c) GVD 曲线

图 2-5 线缺陷三角晶格光子晶体慢光特性随 dy 的变化

水平,数值变化非常缓慢,反映到群速度上可以看出群速度曲线比较平坦, v_g 的取值更加平稳,GVD 值也有所减小,从而使结构获得更好的光传输效率。经上述分析可知,对于平移缺陷行介质柱,左右平移对慢光特性的影响比上下平移对慢光特性的影响弱,色散特性曲线对 dy 的变化与 dx 相比更为敏感,合理地优化这两个结构参数就可得到更好的慢光特性。

2.3 算法的并行性实现

含有缺陷的复杂光子晶体结构的计算往往耗时极长,当计算更加复杂的光子晶体结构时,更能体现出使用并行计算方式提高运算速度的重要性。图 2-6(a)和(b)中二维三角晶格线缺陷光子晶体结构的能带,可以看到仅仅引入一条线缺陷就使得计算量大幅增加,计算时间提高超过 50%。可见在计算复杂的缺陷光子晶体结构时引入并行计算方法是十分必要的,为了得到较为精确的结果,需要设置很大的平面波矢数,运算量极大。将并行算法应用在这些结构中将极大地提升运算效率,减少的运算时间十分可观。

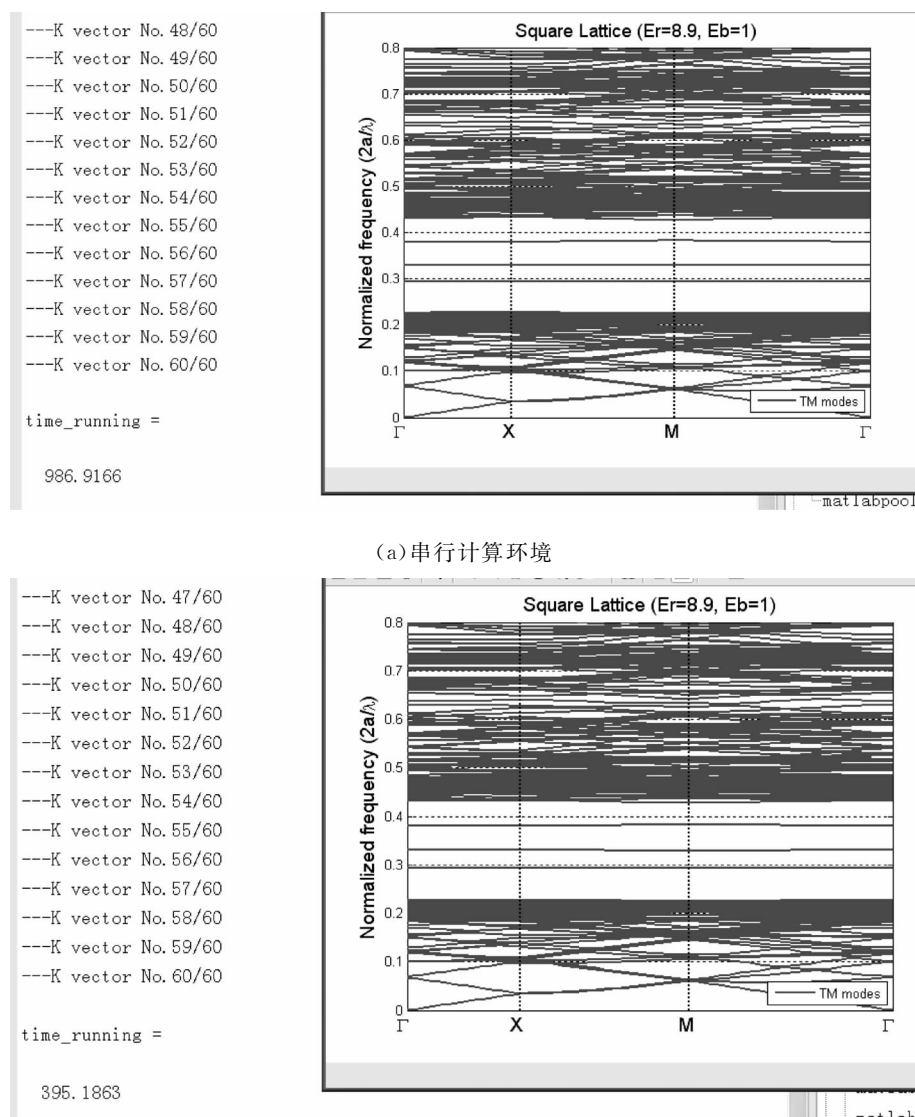
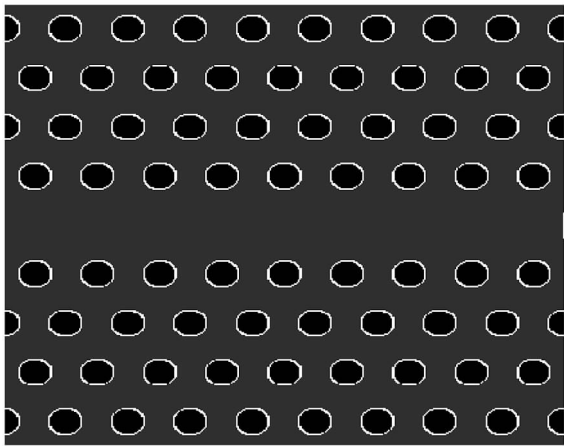


图 2-6 缺陷结构串行与并行计算环境结果对比

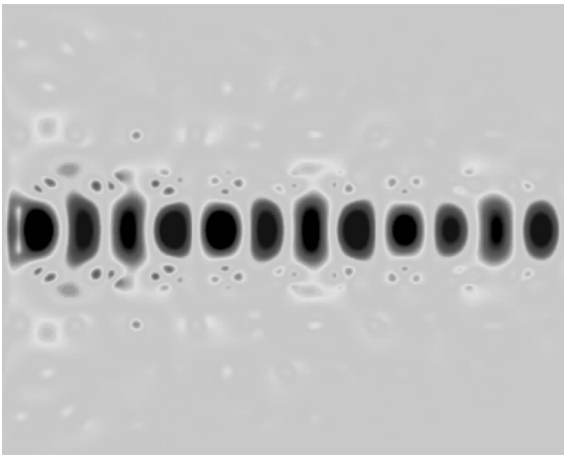
3 二维光子晶体慢光传输特性分析

3.1 光波在缺陷光子晶体结构中的慢光传输

在完整光子晶体中引入缺陷结构后,禁带中会产生缺陷能态,频率在缺陷能态中的光波进入光子晶体后,将被局域在缺陷处无法偏离;如果缺陷类型为线缺陷,将形成光子晶体波导结构,光波将沿缺陷方向传播。



(a)结构平面分布图



(b)传输场分布图

图 3-1 线缺陷三角晶格光子晶体

将完整三角晶格光子晶体中间一行的介质柱移除,得到线缺陷光子晶体波导结构,如图 3-1(a)所示。

将入射光波的频率选取在带隙上,模拟光波在线缺陷波导中的传播,结果如图 3-1(b)所示。在光子晶体中引入线缺陷后,频率在带隙的光波被局限在缺陷处,沿线缺陷防线传播,在其他方向的散射极小,证明该线缺陷结构可以作为光子晶体波导。

3.2 光子晶体透射率的计算

介质柱介电常数为 4.55,衬底材料介电常数为 1,介质柱半径为 0.20a 的线缺陷三角晶格光子晶体结构,计算得到透射率—波长谱如图 3-2 所示。

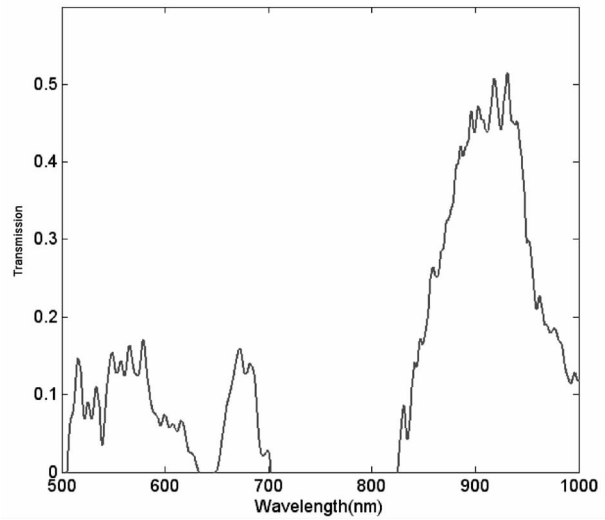


图 3-2 缺陷三角晶格光子晶体结构透射谱

图中结果与文献较为吻合,从理论上分析,当光波的频率落在光子禁带内,则难以通过完整光子晶体,结构透射率较低,而线缺陷光子晶体波导由于带隙中出现了缺陷能态所以透射率较高。从仿真结果上看,完整光子晶体与线缺陷光子晶体相比,带隙的位置出现了偏移,频率位于缺陷能态以外,而落在带隙中的光子在完整光子晶体和线缺陷光子晶体中的透射率均很低。很多文献中均出现类似的结果,这不是由于程序计算误差引起,因为在引入线缺陷后,使结构原有的空间排布及周期性产生了变化,影响了整体能带分布,各能带曲线出现不同程度的偏移进而改变了带隙的位置。

4 耦合腔波导结构与慢光性能分析

4.1 耦合腔波导结构慢光参数优化

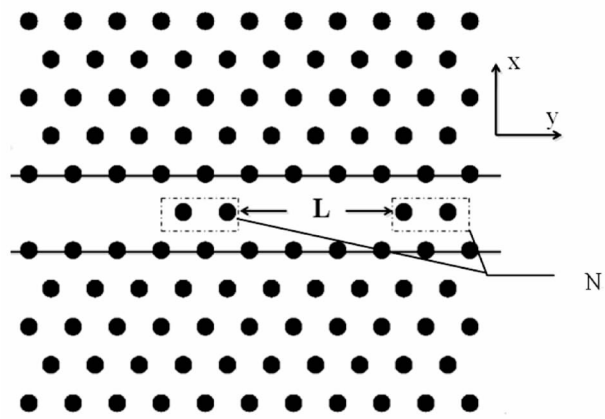


图 4-1 完整三角晶格光子晶体结构透射谱

耦合腔波导结构如图 4-1 所示,介质柱采用介电常数 $\epsilon_a=11$ (近似为硅的介电常数)的材料,晶格常数 $a=0.541\mu m$,介质柱半径 $r=0.4a$,衬底材料为空气($\epsilon_b=1$)。在完整光子晶体的中间引入一条线缺陷,然后在线缺陷中加入等间距排布的介质柱,形成沿水平方向周期排列的相邻微腔结构,每个腔体的长度为 L ,相邻腔体间的介质柱数为 N 。

图 4-2 研究了耦合腔体间周期排布的附加介质柱的数目 N 对波导结构色散特性的影响。当 $N=0$ (即未加入介质柱,仅为线缺陷光子晶体波导)时,色散曲线变化趋势较为剧烈,没有满足实现理想慢光所要求的数值平稳区域;当在线缺陷中附加介质柱形成耦合腔光子晶体波导时,随着所加介质柱数 N 的增加,色散曲线大幅上移,曲线首部与尾部的数值变化非常缓慢。色散曲线中越平坦的部分对应的群速度越小,越陡峭的部分对应的群速度越大。然而若所附加的介质柱数量超过一定限额,色散曲线的平稳区域会发生上翘,又将导致群速度增大及慢光性能的恶化。因此,为使色散曲线出现比较平稳的座椅状能带,实现较好的慢光特性,当附加介质柱数 $N=2$ 时较为合适。

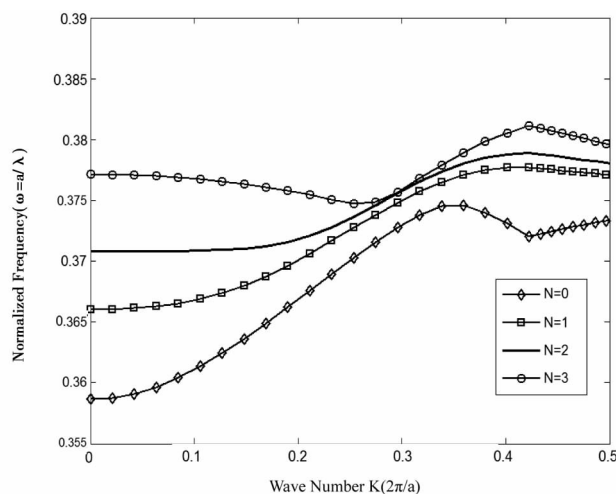
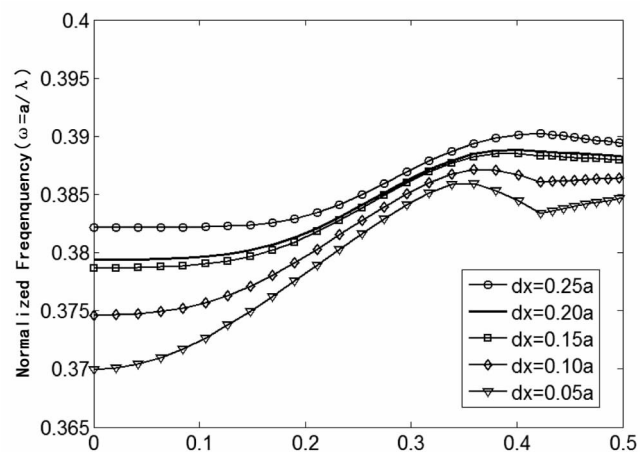
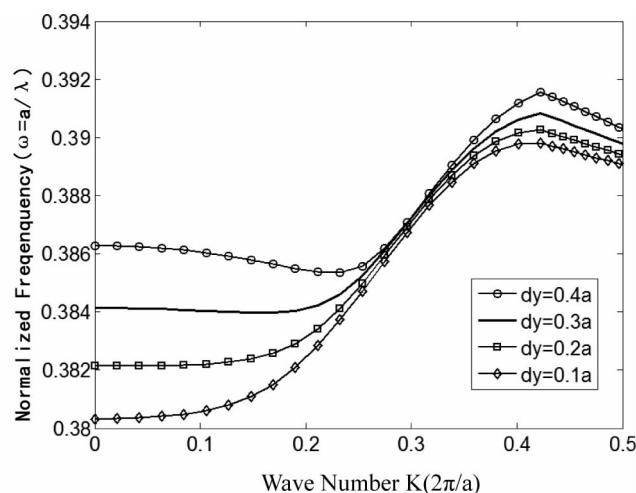


图 4-2 色散特性曲线随腔体间介质柱数量的变化关系

保持 y 方向不变,沿 x 方向平移线缺陷耦合腔上下两侧的介质柱,当平移量 dx 分别取 $0.05a, 0.10a, 0.15a, 0.20a, 0.25a$ 时,如图 4-3(a)所示,随着平移量 dx 的增加,能带逐渐上移,曲线发生弯曲,在 $dx=0.20a$ 时色散曲线表现出较为理想的座椅状能带。接下来保持 $dx=0.20a$,沿 y 方向平移线缺陷耦合腔上下两侧的介质柱,当平移量 dy 分别取 $0.1a, 0.2a, 0.3a, 0.4a, 0.5a$,从图 4-3(b)可以看出,虽然 dy 取值增幅明显高于 dx ,但色散曲线的变化却很小,说明相对于 y 方向介质柱的平移,色散特性对 x 方向介质柱的平移更为敏感。这是因为,沿 x 方向平移介质柱会改变相邻两排介质柱的间距,使整个光子晶体的介质柱密度发生改变,导致光子晶体波导有效折射率增加,从而改善能带色散曲线。



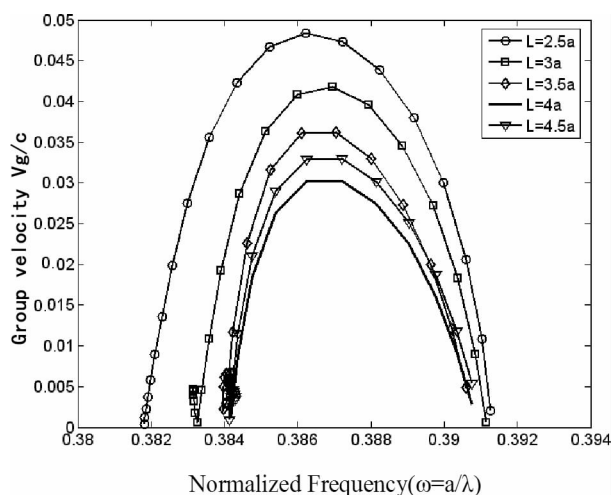
(a) dx 平移量



(b) dy 平移量

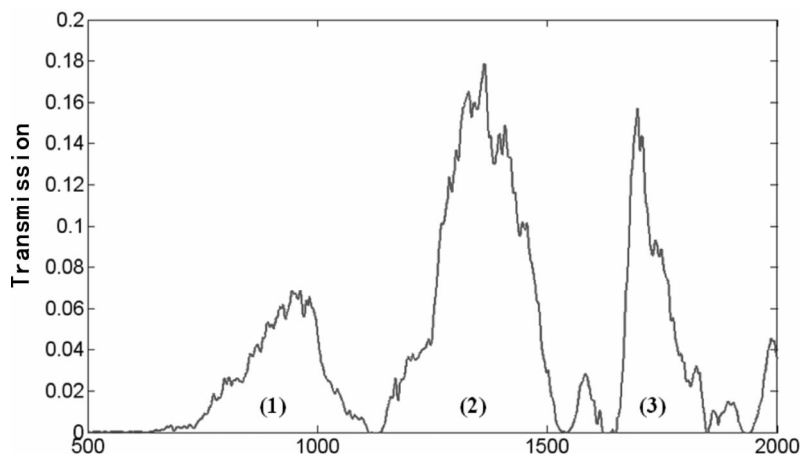
图 4-3 色散特性曲线随不同结构参数的变化关系

至此,完成了光子晶体耦合腔波导色散特性的优化,得到优化的结构参数 $N=2$, $dx=0.2a$, $dy=0.3a$ 。下面在此结构参数下改变腔体的长度 L ,调节耦合腔波导的群速度特性。图 4-4 为群速度曲线随腔体长度 L 的变化,当腔体长度 L 增加时,群速度曲线的峰值先减小后增大,因此选择腔体长度取值为 $L=4a$ 可以获得最小的导模群速度,实现最佳的慢光效果。同时我们发现,在调节腔体长度 L 降低群速度的过程中,有效带宽基本保持不变。而且,经验证,当腔体长度从 $L=2.5a$ 调整到 $L=4.5a$ 的过程中,结构能带色散曲线的形状几乎没有变化。这说明通过改变腔体长度 L ,可以在减小群速度的同时,保持较好的带宽和色散特性。此外由图 4-4 可以看出慢光的有效带宽约 0.006ω ,进一步计算得到波长范围接近 20nm 。

图 4-4 群速度特性曲线随腔体长度 L 的变化关系

4.2 耦合腔波导慢光传输特性验证

为了验证结构具有良好的传输特性,首先作出上述优化结构后的二维三角晶格光子晶体耦合腔波导的透射谱,如图 4-5 所示。

图 4-5 当 $N=2, dx=0.2a, dy=0.3a, L=4a$ 时,耦合腔波导的透射谱

分别选取透射谱中(1)(2)(3)区域中的波长作为入射波长。图 4-6(a)~(d)分别为入射激励光波长 $\lambda=550\text{nm}$, $\lambda=950\text{nm}$, $\lambda=1750\text{nm}$, $\lambda=1450\text{nm}$ 时,光在上述优化结构二维三角晶格光子晶体耦合腔波导中传输的场分布图。从图 4-6(a)可以看出,当入射激励光波长位于图 4-5 所示(1)(2)(3)区域之外时,光在传输过程中由于受到微腔间介质柱的阻拦,产生了极大反射和散射,能量大大损耗,几乎无法通过光子晶体。当入射激励光波长位于图 4-5 所示(1)或(3)区域时,如图 4-6 (b)和图 4-6 (c)所示,虽然光可以通过微腔间介质柱并沿着线缺陷方向传播,但传输过程中会向四周发散,传输损耗仍很大,不利于在实际中应用。而只有当入射激励光波长位于图 4-5 所示(2)区域时,如图 4-6 (d) 所示,光会在散射极小的情况下沿线缺陷方向传播,并无阻碍的通过相邻微腔间的介质柱。证明了该光子晶体波导结构不仅具有良好的慢光特性,而且具有较低的传输损耗。

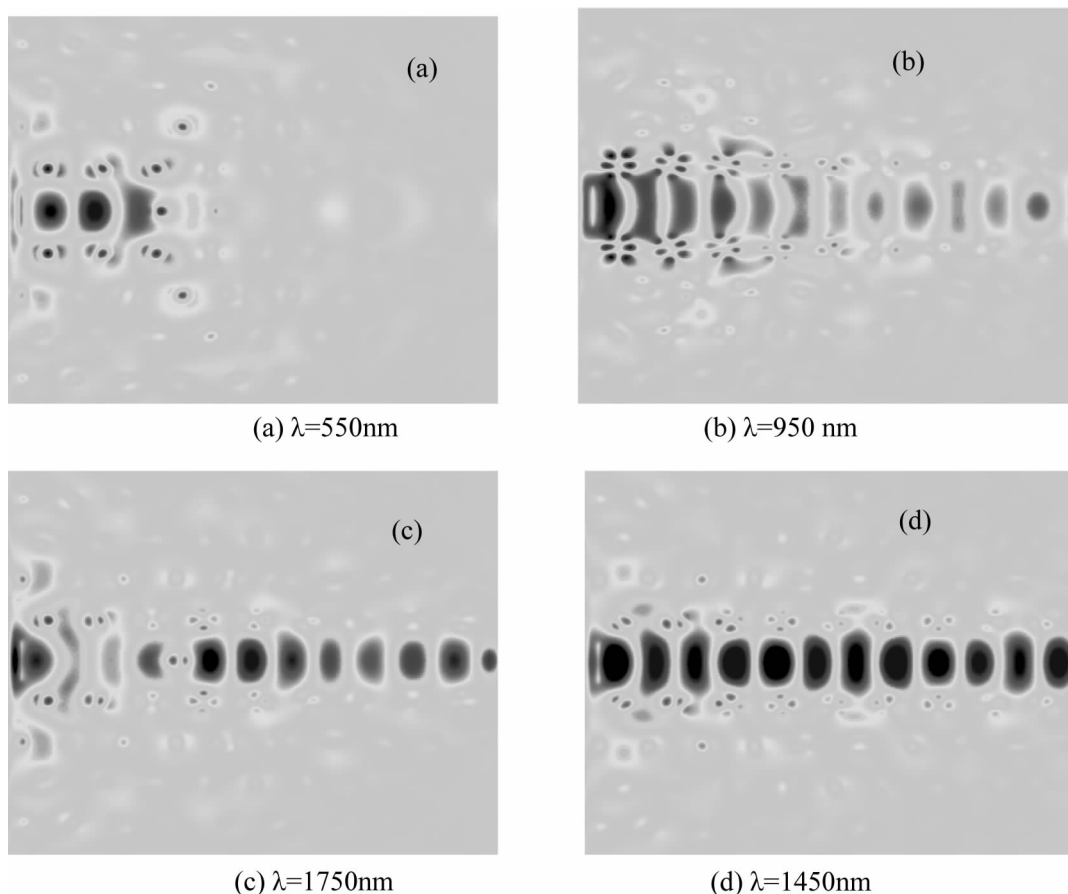


图 4-6 当 $N=2, dx=0.2a, dy=0.3a, L=4a$ 时,耦合腔波导的场分布图

参 考 文 献

- [1] 叶卫民. 光子晶体导论[M]. 北京:科学出版社,2010,1—9
- [2] Joannopoulos J D, Villeneuve P R. Photonic crystals: putting a new twist on light[J]. Nature, 1997, 386:143—149
- [3] 张武生,薛巍. MPI 并行程序设计实例教程[M]. 北京:清华大学出版社,2008,1—15
- [4] 张宁,纪越峰. 光子晶体及其应用研究. 北京联合大学学报(自然科学版)[J]. 2005, vol. 19 No4, 42—45
- [5] 宋清海,徐雷. 光子晶体的原理与应用. 物理实验[J]. 2004, vol. 24, No. 6: 3—7
- [6] 张宁,纪越峰. 光子晶体及其在光通信中的应用研究[J]. 光通信技术, 2004, vol. 7 :4—8
- [7] 于学亚,曾兆华,杨建文,陈用烈. 光子晶体的研究现状与最新进展[J]. 人工晶体学报, 2002, 33(6): 602—607
- [8] 韩利红,刘立明,郭璇,等. 正方形空气孔 L3 型光子晶体平板微腔的研究与设计[J]. 光学学报, 2013, 33(11): 1116005
- [9] X Yang, A Ishikawa, X Yin, et al. . Hybrid photonic plasmonic crystal nanocavities[J]. ACS Nano, 2011, 5 (4): 2831—2838
- [10] P Xu, K Yao, J Zheng, et al. . Slotted photonic crystal nanobeam cavity with parabolic modulated width stack for refractive index sensing[J]. Opt. Express, 2013, 21(22), 26908—26913
- [11] P Xu, Q Huang, et al. . Silicon hybrid plasmonic bragg grating reflectors and high Q-factor micro-cavities[J].

- Optics Communications, 2013, 289: 81–84
- [12] R M Ma, R F Oulton, et al. . Room-temperature sub-diffraction-limited Plasmon laser by total internal reflection [J]. Nat Mater, 2011, 10(2): 110–113
- [13] Luo Xin, Zou Xihua, et al. . Narrow-band filter of surface Plasmon based on dual-section metal-insulator-metal structure[J]. Acta Optica Sinica, 2013, 33(11): 1123003
- [14] V Georgios, Zhangfu Yu, et al. . Metal-dielectric-metal plasmonic waveguide devices for manipulating light at the nanoscale[J]. Chin Opt Lett, 2009, 7(4): 302–308
- [15] Feng Chen, Feng Guoying, et al. . Characteristic analysis on photonic crystal laser cavity with one-dimensional photonic bandgap[J]. Chinese J Lasers, 2012, 39(8): 0802009
- [16] Fleming J G, Lin S Y, et al. All-metallic three-dimensional photonic crystals with a large infrared bandgap. Nature, 2012, 417: 53–55
- [17] 毛宇, 童元伟. 改变光子晶体表面结构对负折射透射光强的影响[J]. 光子学报, 2010, 39(9): 1562–1566
- [18] Lodahl P, et al. Controlling the dynamics of spontaneous emission from quantum dots by photonic crystal[J]. Nature, 2004, 430: 654–657
- [19] Painter O, Lee R K, et al. Two-dimensional photonic band-gap defect mode laser[J]. Science, 1999, 284: 1819–1821
- [20] 万钧, 张淳, 王灵俊, 资剑. 光子晶体及其应用[J]. 物理学报, 1999, 28(7): 393–398
- [21] 詹仪, 郑义, 徐云峰. 二维介质柱光子晶体波导的优化设计[J]. 光子学报, 2008, 37(5): 1130–1133
- [22] 史军, 刘勇科, 林旭彪等. 一种衍生于三角晶格的二维光子晶体[J]. 光子学报, 2008, 37(5): 940–943
- [23] 胡应涛, 李运涛, 李智勇等. 硅基波导慢光器件及其应用[J]. 物理学和高新技术, 2010, 39(4): 267–271
- [24] 张栋, 赵建林, 吕淑媛. 低群速度色散和低损耗的二维光子晶体慢光波导[J]. 光学学报, 2011, 31(1): 1–5
- [25] 杨波, 梁静秋, 梁中翥, 崔乃迪, 周建伟, 宁永强, 王维彪. 二维金属型光子晶体带隙研究[J]. 激光与红外, 2011, 41(3): 314–318
- [26] Min Qiu, Sailing He. A nonorthogonal finite difference time domain method for computing the band structure of a two-dimensional photonic crystal with dielectric and metallic inclusions[J]. Journal of Applied Physics, 2000, 87(12): 8268–8275
- [27] 张玉. 电磁场并行计算[M]. 西安: 西安电子科技大学出版社, 2005, 5–27
- [28] 肖国宏, 傅君眉, 冯恩信. 具有缺陷结构的三角形二维光子晶体结构参量对本征模的影响[J]. 光子学报, 2008, 37(5): 725–728
- [29] 杜晓宇, 郑婉华, 任刚等. 二维光子晶体耦合腔阵列的慢波效应研究[J]. 物理学报, 2008, 57(1): 571–575
- [30] 吕淑媛, 赵建林, 张栋. 二维正方光子晶体波导中的慢光传输[J]. 光子学报, 2009, 38(9): 2224–2228
- [31] 殷建玲, 黄旭光, 刘颂豪等. 介质柱形状对准晶光子晶体带隙特性的影响[J]. 中国激光, 2010, 39(2): 567–571
- [32] 曲连杰, 杨越德, 黄永箴. 光子晶体波导慢光特性研究[J]. 光学学报, 2011, 31(1): 0113002
- [33] 张佑专, 李志荣, 郑咏梅, 王京霞, 宋延林, 江雷. 仿生制备功能性聚合物光子晶体[J]. 高分子学报, 2010(11)
- [34] 邓开发, 是度芳, 蒋美萍, 李承芳. 光子晶体研究进展量子电子学报[J], 2004, 21(5): 555–563
- [35] 李正华, 薛燕陵, 沈廷根. 光子晶体内的慢光及其应用[J]. 激光与红外, 2008, 38(1): 4–6
- [36] 张友俊, 杨庆祥, 李英. 光子晶体发展和应用[J]. 上海大学学报(自然科学版), 2004. 04, 10(3): 283–288
- [37] 倪培根. 光子晶体制备技术和应用研究进展[J]. 物理学报, 2010, 59(1): 340–349
- [38] 李燕, 谢娟, 邓宏, 徐自强. 光子晶体的研究进展[J]. 材料导报, 2006. 02, 20(1): 10–12
- [39] 李会玲, 王京霞, 宋延林. 光子晶体的制备与应用研究[J]. 自然杂志, 2009, 31(3): 154–157
- [40] 李欣, 杨明, 郭士亮, 李志全. 含随机介质的光子晶体波导特性[J]. 中国激光, 2012(10)
- [41] 鲁辉, 田慧平, 李长红, 纪越峰. 基于二维光子晶体耦合腔波导的新型慢光结构研究[J]. 物理学报, 2009. 03, 58(3): 2049–2055

- [42] 李长红,田慧平,鲁辉,纪越峰,波导基本结构对光子晶体耦合腔光波导慢光特性的影响[J],光学学报,2010,30(7):2109—2114
- [43] 徐小果. 基于 FDTD 法的二维光子晶体传输特性研究[D]. 成都:西南交通大学通信与信息系统系,2009
- [44] 冯琛,冯国英,周昊,陈念江,周寿桓. 一维光子带隙光子晶体激光腔的特性分析[J]. 中国激光. 2012(08)
- [45] 丁巨超. 光子晶体带隙时域仿真方法的研究[D]. 天津:天津大学精密仪器与光电子工程学院,2006
- [46] 葛德彪 闫玉波. 电磁波时域有限差分法(第二版)[M]. 西安:西安电子科技大学出版社,2005,27—120
- [47] Nobuyuki Matsuda, et al. Slow light enhanced correlated photon pair generation in photonic-crystal coupled-resonator optical waveguides[J]. Opt. Express, 2013, 21(7), 8596—8604
- [48] Huang Chongqing, et al. The FDTD Method in Electromagnetic Field[J]. Journal of Yueyang Normal University (Natural Sciences), 2003, 16(1):75—78

Slow light properties research in photonic crystals

Quan Wei

(School of Electronic and Information Engineering, North China Univ. of Tech. , Beijing, 100144)

Abstract In this thesis, properties of two-dimensional photonic crystal band-gap, waveguide slow light and transmission characteristic is analyzed. Band gap characteristics of two-dimensional hexagon column photonic crystals are analyzed with numerical computation method. The slow light properties of defected photonic crystal waveguide and the influence of defect dielectric column are analyzed by calculating the photonic band, the group velocity and the group velocity dispersion of photonic crystal structures. Then the transfer characteristic of light wave whose frequency is in or out of the photonic band gap is studied and transmission spectrums of different structures are given. Finally, the slow light property of coupled-cavity waveguide is analyzed and a good transmission character of the proposed structure is verified by researching transmission field distribution under different incident wavelength in photonic crystal coupled cavity waveguides.

Key words photonic crystal; filter; slow light; FDTD; PWE

分布式生物特征识别技术研究

研究生 于晓婕

导师 王一丁

(北方工业大学电子信息工程学院 北京 100144)

摘 要 本文首先对图像采集、图像预处理方法进行实验分析,并针对分布式系统和手背静脉图像的特点,提出基于小波的多尺度局部二值模式(Local binary patterns, LBP)算子的权重融合方法。通过基础研究,最终设计一套分布式手背静脉识别门禁系统。

关 键 词 手背静脉识别;生物特征识别;多尺度 LBP;分布式系统

作者简介 于晓婕,女,1988 年出生,硕士研究生,主要研究方向为图像处理与模式识别。

E-mail: gengyuan927@163.com

1 绪论

运用手背静脉信息进行身份识别是近些年发展起来的生物特征识别方法^[1-5],主要根据手背皮下的血管脉络提取特征信息。手背静脉识别技术具有其独特的优势:活体识别,保证用户的静脉特征不容易被伪造;内部特征,不宜被污染或破坏;非接触式,杜绝用户恐惧心理;对光照条件不敏感,由于使用近红外设备采集,因而相比普通的成像设备,对光照变化不敏感。因而,基于手背静脉识别具备大量优势,本文目的是设计分布式手背静脉识别系统。

2 图像采集与数据库建立

近红外手背静脉图像的采集设备内、外部图如图 2.1 所示。

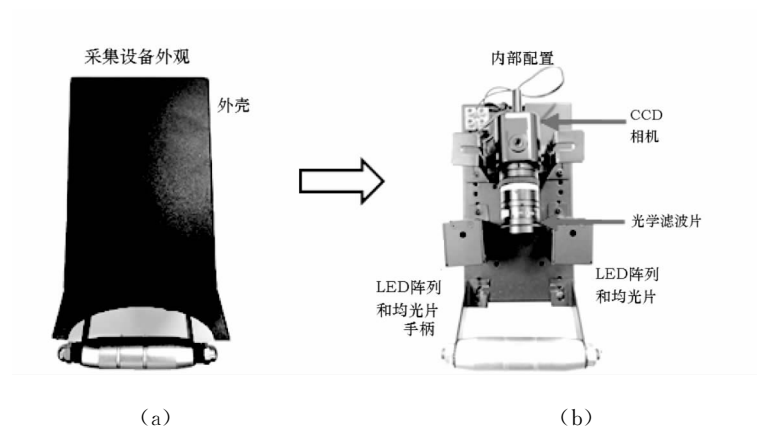


图 2.1 (a)采集设备内部结构 (b)采集设备外部结构

采集手背静脉图像并构建图像数据库。随机邀请志愿者紧握手柄 20 次,左右手交替采集,即共 20 个人,每个手采集 10 次。最终共得到 2040 张手背静脉图像,并且组建成一数据库,该库中存储的静脉图像分辨率统一是 640×480 像素。

3 手背静脉图像处理

分布式系统,不同终端的采集设备对后续的认识会产生非常大的影响。采集设备在设计制造上本身可能存在差异,而且采集设备所处的光照环境也各有差异,以及采集手背者的姿势不同,这些因素都会导致即使相同的手背采集到的手背静脉图像都有较大差异,这些因素对后续的特征提取和识别有较大影响,不利于实现分布式的手背静脉系统。为了克服这些现象,并针对采集设备设计和实现的具体情况,以及考虑分布式系统的识别率和识别速度,特征提取工作之前,首先要对手背静脉图像进行预处理,根据实验分析,预处理工作中主要含有图像校正,感兴趣区提取,噪声去除,图像归一化等操作。最终得到一致性较高的图像,为后续识别系统设计起到良好铺垫作用。

4 正文图像采集与数据库建立

4.1 基于 LBP 的小波分解

对于旋转不变一致性模式的 LBP 直方图,使得手背静脉图像局部灰度变化表现的更加紧凑。但是另一方面,由于使用更短的特征向量来表示静脉图像特征,缩减了特征表现的能力,限制了更强的鉴别能力影响获取更高识别率。

多尺度 LBP 算法采用的就是利用算子不同尺度的半径来提取特征信息,使得能够在更大范围的邻域中获取灰度变化的信息^[6],或者采用图像下采样的方法得到不同尺度的图像,这样使得能够捕获到不同分辨率的图像信息^[7]。这种方法对噪声的影响不敏感,而且不会引起混淆现象,因为它在下采样前涉及到运用低通滤波进行处理。

采用根据小波变换的分解方法一方面使得图像下采样,另一方面又获取图像的方向信息。所以提出结合分块 LBP 与不同的采样半径来得到综合性的图像特征描述算子。图像的离散小波变换可以实现为两个一维小波变换过程。通过使用两个不同频带的滤波器组包括一个高通滤波和一个低通滤波,如图 4.1 所示。

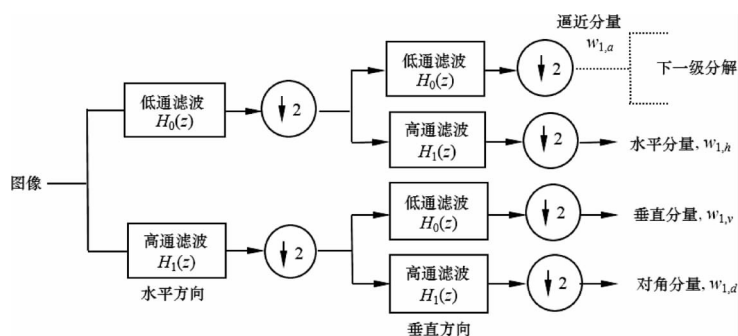


图 4.1 基于小波的图像分解原理

二级 Haar 小波变换,输出的小波分量各部分如图 4.2 所示。从图 4.2 中可以看出,静脉的所有结构信息分别在小波分解后的两级逼近图像中都非常清晰。对比与水平方向的小波分量,垂直方向的小波分量含有更多的静脉结构的有用信息,这也反映出更多的静脉主线处在主要方向是垂直的方向。对角的小波分量完全看不清静脉的结构,几乎不含有有用的信息。

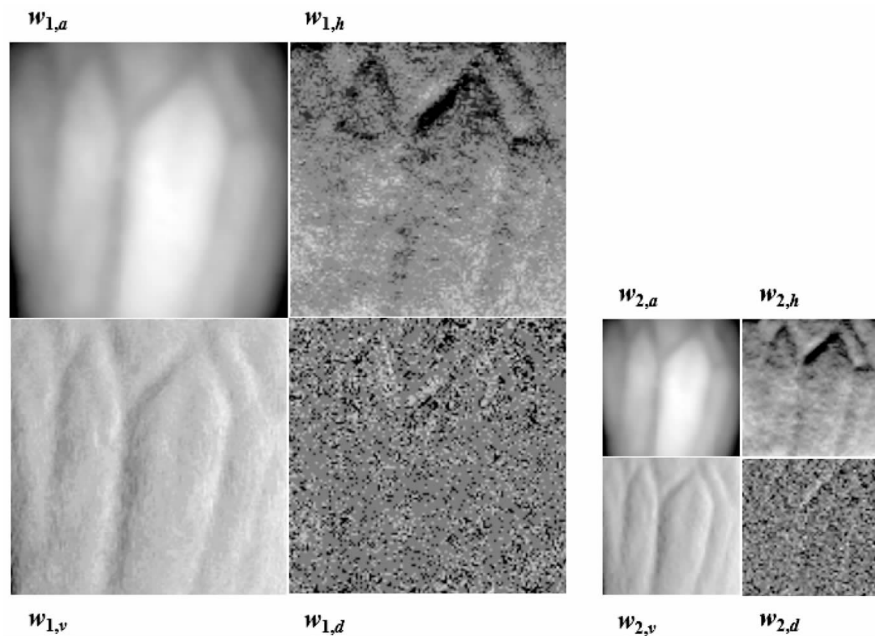


图 4.2 基于小波的图像分解

4.2 实验结果与分析

为了进一步验证提出算子的鲁棒性,该组试验比较的是提出的算子与其它基于 LBP 的算子的识别率的情况。这些 LBP 算子包括原始的 LBP(Original LBP),中心对称 LBP(CS-LBP)^[8],主 LBP(DLBP)^[9],改进 LBP(ILBP)^[10],中值 LBP(MLBP)^[11],传统多尺度 LBP(Traditional multi-scale LBP)^[12]。

表 4.1 为各个算子的识别率结果,从表中可知,新提出的算子(Proposed method)通过采用多分辨率灰度级,位置以及方向信息,结合多方面的优势最终使得识别率达 99.12%。

表 4.1 不同算子的识别率

算子	识别率
Original LBP	95.49%
CS-LBP	97.75%
DLTP	89.71%
ILBP	95.10%
MBP	94.61%
Traditional multi-scale LBP	97.84%
Proposed method	99.12%

在众多的算子中,提出的算子和传统的多尺度算子有相对较高的识别率,而且新提出的算子识别率更高。图 4.3 中基于不同的训练集和测试集进行十次重复的实验得到最大识别率、最小识别率以及平均识别率。图 4.3(a)中为传统多尺度 LBP 算子由 $LBP_{8,2}^{rU2}$, $LBP_{8,3}^{rU2}$, $LBP_{8,4}^{rU2}$ 的直方图融合拼接而成的识别率。图 4.3(b)中为提出的基于权重融合的小波变换多尺度算子的识别率。虽然传统多尺度算子与提出的多尺度算子识别率都是随着训练图片的增加,两者识别率有相同的增长趋势和减小的变化,但是新提出的多尺度算子拥有更好的识别率以及更小的识别率变化。进一步可以说明新提出的多尺度算子有更强的鲁棒性。根据实验分析和统计,基于识别的性能和识别率的变化,分析可知在实验时至少训练集图像为 5 张图像时能得到较为满意的识别结果。接下来试验中采用训练集和测试集分别为每类手背静脉图像中的 5 张来进行实验分析。

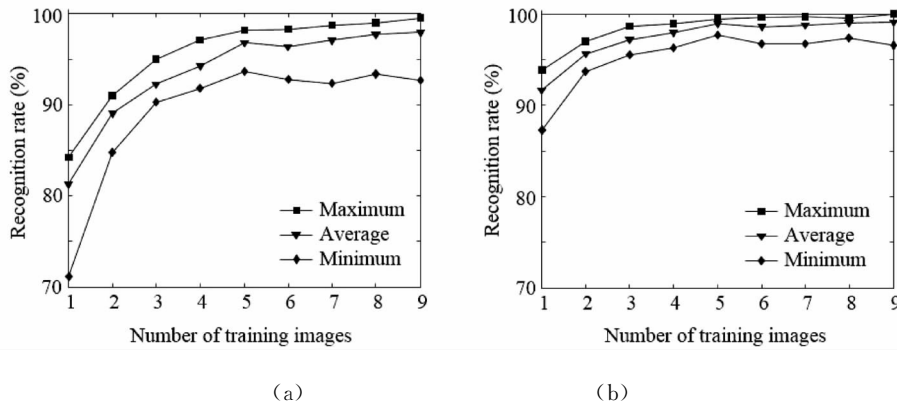


图 4.3 各个训练集得到最大、平均以及最小识别率变化曲线情况
(a)传统多尺度 LBP (b)基于小波变换的多尺度 LBP

实验分别得到传统多尺度算子以及新提出多尺度算子的错误接受率(False Acceptance Rate, FAR)以及错误拒接率(False Rejection Rate, FRR)的曲线并得到识别阈值大小,如图 4.4(a)和 4.5(a)。阈值设置是在 0 到 1 之间且步长为 0.01,提出的特征算子当阈值为 0.713 时,等错误概率(Equal Error Rate, EER)为 1.79%。而相比于,传统多尺度算子阈值为 0.746 时,EER 为 3.92%,提出的特征算子的 EER 更小,体现出其良好的整体性能。图 4.4(b)和 4.5(b)分别为传统多尺度算子和新提出的特征算子的受试者工作特性曲线(Receiver Operating Characteristic Curve, ROC)。

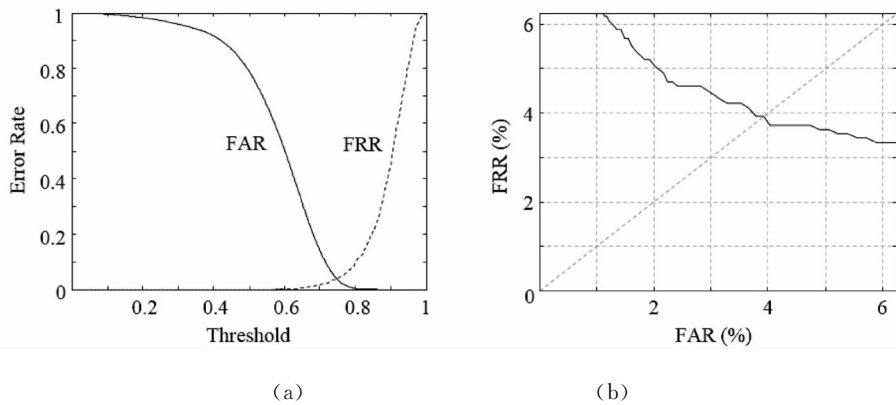


图 4.4 传统多尺度 LBP (a)FAR 和 FRR 曲线 (b)ROC 曲线

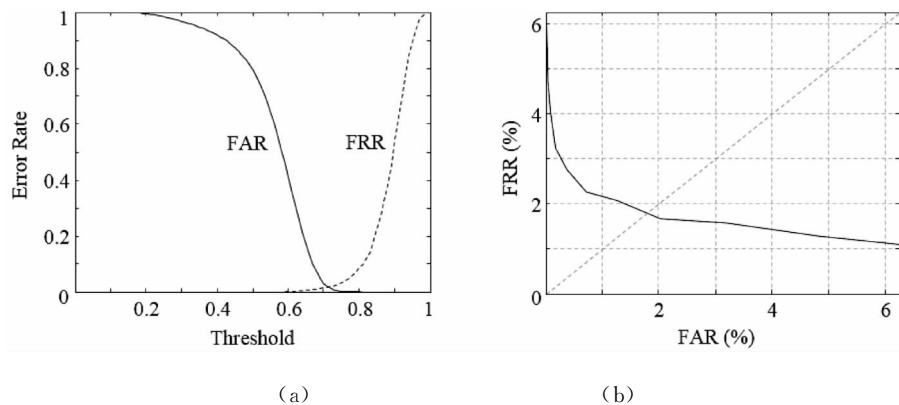


图 4.5 提出的多尺度 LBP (a)FAR 和 FRR 曲线 (b)ROC 曲线

5 分布式手背静脉系统研究

5.1 系统介绍

通过前几章的基础研究工作,了解构建生物特征识别系统的整体流程和每一阶段的工作。通过研制一套基于近红外的图像设备,进行手背静脉图像生物采集,分析实际环境以及使用方式,实验得到有效的图像预处理方法,得到一致性较高图像使得构建分布式系统尽可能减小采集图像带来的干扰,为系统能够得到鲁棒特征做好基础工作。经过试验分析提出的分布式的手背静脉门禁系统,具有可靠识别、方便使用、高安全性、快速识别、低成本以及方便安装等一系列特点。

如图 5.1 所示,本系统为了低成本且便于安装,是基于简易化的硬件设备设计的。系统包含有微型电脑通过两个 USB 接口分别连接着手背静脉图像采集设备和继电器,且有通过 VGA 线连接着显

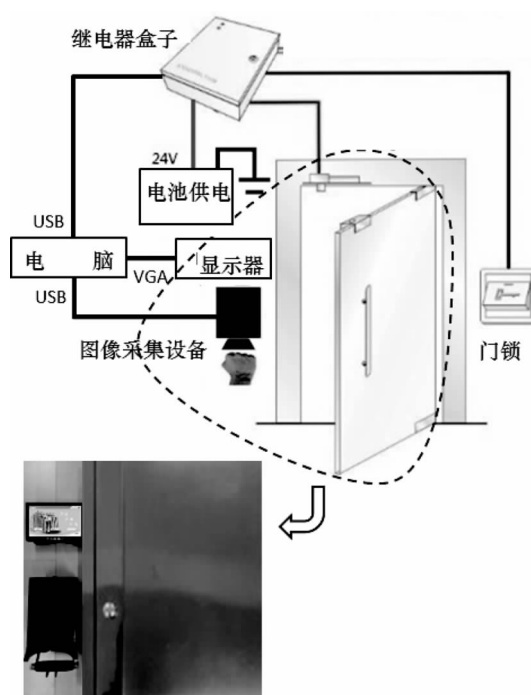


图 5.1 系统硬件组成图

示器。黑色外壳的采集设备是用于静脉图像的采集,而显示器又能给用户呈现良好的视觉效果。继电器主要是用于支配电流的通断,进而再一步支配电磁锁的吸合。电磁铁分别安装在门框上和门上,通电时强烈的磁铁作用使门保持锁住。当继电器使得电池锁保持吸合状态,经授权的用户的手背静脉特征识别成功后,信号使得继电器断开从而电磁锁也断开使得门能打开。

图 5.2 为试验自制设计的分布式手背静脉门禁系统的模拟实物图。主要由采集模块、电脑显示器模块、电源、继电器模块、门禁模块组成。这些模块一起构成一套简易的终端设备。

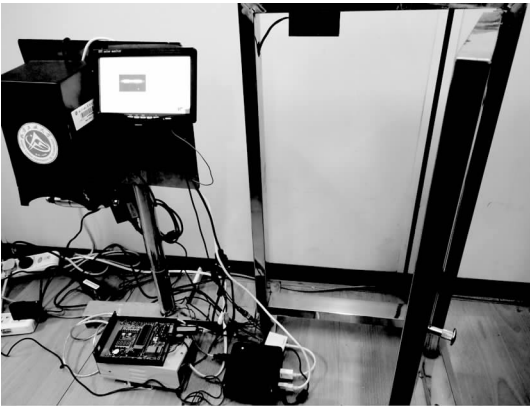


图 5.2 系统设计搭建的终端模拟实物图

图 5.3 为系统软件算法流程图,系统由程序控制,软件分为两模块。其中一模块为系统注册端,用于提取特征并发送每一用户的特征模板存储于数据库。系统另一模块为识别端,当手背放入采集设备,把提取到手背静脉图像特征同库中的特征匹配,寻找最佳类。

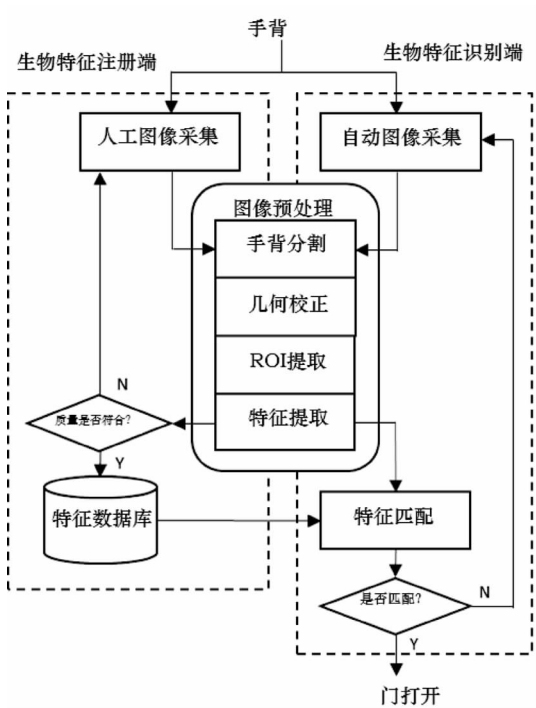


图 5.3 系统软件算法流程图

当用户手放入采集设备中并紧握把手,注册端手动的被激活进行采集图像,而在识别端在手背放入采集设备后就自动每隔一秒进行采集图像,并处理后进行识别。注册端和识别端从手背图像的分割到特征的提取都采用相同的图像预处理方法。对于特征提取,提取完毕的特征在注册端会首先会被检测该特征是否符合要求,若符合要求就会存储到数据库中,否则会提示重新进行采集图像。

5.2 系统介绍

5.2.1 系统架构设计

基于分布式手背静脉识别系统是生物识别与网络化发展带来的必然产物。在构建分布式的手背静脉识别系统中采用网络化的技术,使得拥有分布式的多个终端设备,能处于任意的位置。服务器能够集中管理多个终端点,而且网络连接着各个终端与服务器,能够方便的增添终端设备。分布式的架构使得系统实现网络化的交互以及使的系统结构更加灵活,终端设备散布更加广泛。多个终端设备共同由相同的服务器进行分配管理,使的数据库中拥有更多的采集数据,扩大了特征库数据。基于分布式手背静脉识别门禁系统硬件架构设计可由图 5.4 直观清晰体现。

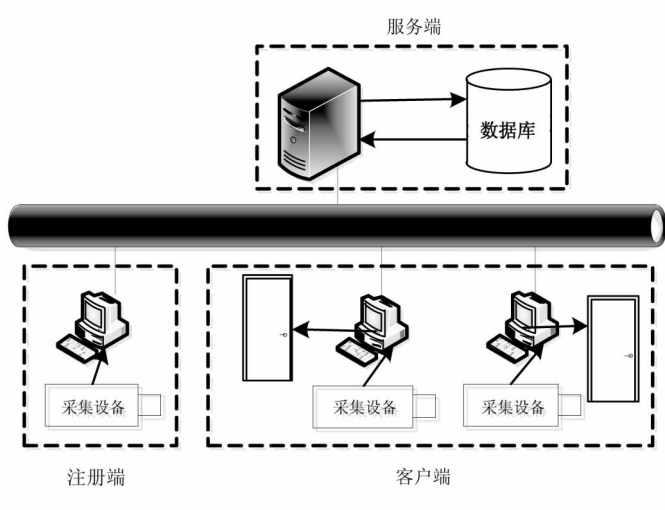


图 5.4 系统硬件架构

分布式系统用户终端能够共享服务端中数据库信息,且系统中各个客户端凭借网络通讯与服务端持续不断地交换信息。分布式系统终端设备能够同时并行连接服务端,服务端得到连接反馈信息后,统一由服务端执行相应的客户端请求。各个终端分别进行相应的图像预处理与特征提取操作,终端承担这些操作后,减轻了服务器端负担,使得服务端能够快速进行识别匹配任务。各个用户终端能够并发的发送请求信息,服务端能够同步的处理这些请求,每一客户端连接成功服务端后,服务端都会新开辟一个线程,专门服务于此终端。服务器端根据相应的机制处理数据并返回处理信息结果。

网络间通讯机制主要根据事先制定的网络通讯协议进行信息传输交换,而于本文分布式识别系统中通讯协议采用的是 TCP/IP 协议并采用 socket 通讯机制。Socket 又称为套接字,为通讯链路中句柄,并可以用其描述 IP 地址以及端口号信息。系统应用软件程序一般使用套接字来给网络发出申请或者反馈并应答网络,本系统 CSocketServer 类用于服务器端,CSocketClient 类则用于客户端,而 Socket 套接字是于网络构建连接时需要使用的。本系统服务端程序由多个线程构成,其中一线程主要管理服务端中连接请求的任务,服务端会一直处于等待连接,实时监视着网络请求,并允许许多客户端

连接。当用户在客户端发出建立连接申请时,客户端套接字最开始会描述要建立连接的相应服务端 IP 以及端口号信息,接着会向服务器发送建立连接的申请。一旦服务端监测到或者接收到客户端要构建连接申请,就会立刻反馈给客户端的申请,并且立即建立新通讯线程,把服务端的 IP 和端口号信息又发送到客户端。若客户端确认,那么连接就能够成功完成,而后服务端又会接着处于监听并等待状态,等待其它客户端构建连接申请。TCP/IP 协议是运用较为广泛的一种面向连接端到端协定,并且是可靠传输方式,能够适合大量信息传递环境,在本文分布式手背静脉门禁系统中采用的就是可靠的 TCP/IP 协议。

5.2.2 系统软件设计

本章系统软件是基于 Windows 平台和 MFC 框架进行构建研发的,其中编程开发环境主要运用 Microsoft visual studio 2008,程序语言主要是 C++^[13],后台数据库采用 Microsoft SQL server 2008,运用主要的开源计算机视觉库是 Opencv 2.10。

本章构建的系统软件主要有客户端和服务端两个部分。在客户端软件中又包括两个模块,一是注册端,而另一是识别端。这两个模块设计实用且操作简便,用户界面可分别如图 5.5 和 5.6。

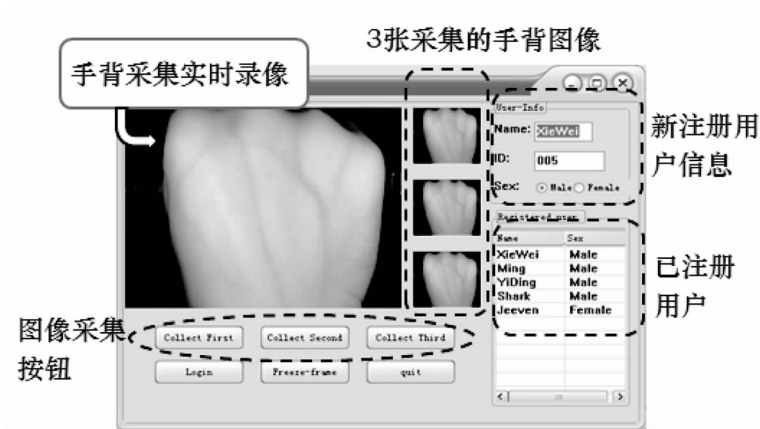


图 5.5 注册端软件界面



图 5.6 识别端软件界面

用户注册需要采集三张图像,这三图像分块后各自串联得到的相应的 LBP 直方图,并要求足够的相似,如果采集手背图像的三个相似度值都大于该阈值,则把三张图像的串联直方图的平均值存储到数据库中作为新用户的注册模板。在识别端,手背放入采集设备后系统连续的采集图像进行识别,识别成功则提示是欢迎用户进入,若不能匹配成功会提示错误信息,提醒用户调整手背姿势重新进行识别。如图 5.6,识别端系统软件还会提供日志信息包括进入用户的用户名以及进入的时间,能够灵活的作为考勤软件使用。

服务端主要功能是监听客户端请求,以及进行识别匹配工作后返回给客户端结果。服务端软件需要无间断的保持运行状态,当一有请求时立刻进行相应工作,而且服务端并不需要较多的界面操作,只需在后台进行处理,所以服务端软件界面简单,如图 5.7 所示。服务端记录下发出连接请求的客户端 ip 和端口号,以及终端结束工作的状态,最终生成日志文件以供查看。



图 5.7 服务端软件界面

6 分布式手背静脉系统研究

本章从硬件的设计到软件的实现介绍了完整的基于分布式手背静脉门禁系统的实现过程。使用近红外设备采集手背图像,有效的算法进行图像处理,提出的方法便捷且具有高可靠性,又由于手背静脉识别固有的优势,这套分布式手背静脉门禁系统能够达到较为满意的效果。这套系统拥有仅为 0.02% 的 EER,基于数据库实验分析得出的鲁棒处理参数,且系统计算复杂度低使得快速的反应时间仅有大约 1s,而且使用现成的硬件设备和较少的部件使得整套系统安装简单且成本低。分布式的架构设计,使得系统能够拥有多客户端,能在多地共同使用,使用更加方便灵活、且系统扩展性强。

参 考 文 献

- [1] Wang K J, Zhang Y, Yuan Z, et al. Hand vein recognition based on multi supplemental features of multi-classifier fusion decision[C]. Proceedings of the 2006 IEEE International Conference on Mechatronics and Automation, Luoyang: 2006, 1790—1795

-
- [2] Wang L, Leedham G, Cho S Y. Infrared imaging of hand vein patterns for biometric purposes [J]. IET Computer Vision, 2007, 1(3-4): 113-122
 - [3] Wu X Q, Gao E Y, Tang Y B, et al. A novel biometric system based on hand vein[C]. Proceedings of 5th International Conference on Frontier of Computer Science and Technology. Changchun: 2010, 522-526
 - [4] Yüksel A, AKarun L, Sankur B. Biometric identification through hand vein patterns[C]. 2010 IEEE 18th Signal Processing and Communications Applications Conference. Diyarbakir: 2010, 708-711
 - [5] 燕青宇. 手背静脉身份识别算法研究[学位论文]. 北京:北方工业大学, 2012
 - [6] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987
 - [7] Wang Y, Yan Q, Li K. Hand vein recognition based on multi-scale LBP and wavelet[C]. Proceedings of 2011 International Conference on Wavelet Analysis and Pattern Recognition: 2011, 214-218
 - [8] Heikkila M, Pietikainen M, Schmid C. Description of interest regions with local binary patterns [J]. Pattern Recognition, 2009, 42(3): 425-436
 - [9] Bendada A, Akhloufi M. Multispectral face recognition in texture space [C]. 7th Canadian Conference on Computer and Robot Vision: 2010, 101-106
 - [10] Jin H, Liu Q, Lu H. Face detection using improved LBP under Bayesian framework [C], Proceedings of Third International Conference on Image and Graphics: 2004, 306-309
 - [11] Adel H, Guna S, Kannappan P. Rotationally invariant hashing of median binary patterns for texture classification [C]. 5th International Conference on Image Analysis and Recognition: 2008, 619-629
 - [12] Guo Z, Lei Z, David Z. Hierarchical multiscale LBP for face and palmprint recognition[C]. 17th IEEE International Conference on Image Processing: 2010, 4521-4524
 - [13] 张铮,王艳平,薛桂香. 数字图像处理与机器视觉—Visual C++与 Matlab 实现[M]. 北京:人民邮电出版社, 2010

Research on Distributed Biometric Identification Technology

Yu Xiaojie

(School of Electronic and Information Engineering, North China Univ. of Tech. , Beijing, 100144)

Abstract The paper focused on experimental analysis of the image acquisition and image preprocessing at first. Considering the characteristics of distributed system and hand-dorsa vein image, a novel method called weighted fusion of wavelet-based multi-scale LBP is proposed. Through the basic research, finally distributed hand-dorsa vein recognition entrance guard system is designed.

Key words hand dorsal vein recognition; biometric features recognition; multi-scale LBP; distributed system

基于光通信的 LDPC 通用编解码器的实现与研究

研究生 张 锋

导师 刘文楷

(北方工业大学信息工程学院 北京 100144)

摘 要 在紫外光(UV)通信系统中,由于噪声对信号产生明显的干扰。为了改善系统的紫外光的通信质量,本文实现了一种改进的低密度奇偶校验码(LDPC)算法并仿真了该算法在紫外光信道模型中非视距传播(NLOS)的性能。设计了 LDPC(960,480)和 RS(18,10)编解码器平台,实验结果表明:在给定发送功率和 10^{-6} 误码率(BER)情况下,相对于未编码的系统,RS 编解码仅增加了 42% 的可靠通信距离,采用 LDPC 编解码却能够增加 100% 左右的通信距离,有效改善系统的通信性能。

关键词 紫外光通信; 低密度奇偶校验码; 里所编码; 非视距传播; 误码率

作者简介 张锋,男,1988 年出生,硕士研究生,主要研究方向为无线光通信。

E-mail:oppositewind@163.com

1 绪论

紫外光通信是基于大气散射和吸收的无线光通信技术^[1]。大气对短波长的紫外光具有很强的吸附和散射作用^[2]。紫外光的波长范围为 200—280nm,其比可见光中紫光的频谱位置还要高一些而得名。大气中有一层保护人类的免受紫外线辐射的臭氧层存在,其对 UV 的吸收作用使得地球表面的紫外光辐射非常的小,这形成所谓的“日盲区^[3]”,紫外光的减少使得地面上使用 UV 进行通信的时候环境背景的干扰就小,对 UV 通信越有利。同时,大气的散射作用使紫外光的传输方向不再是单一的方向,而是向各个方向发散,利用其作为通信的载体成为可能。但是由于大气吸附作用导致紫外光不能传输很远,因此只适合短距离区域通信。

由于实际通信中,终端分散区域比较大,而紫外光通信由于空气的散热和吸附作用,当距离比较远的时候,通信的误码率必然极高。解决这一问题最好的方法是引入传统通信中的信道编码,这样即便距离较远,信道编码的纠错能力也可以保证信息的顺畅传递。

LDPC 码作为最接近香农极限的编解码方案^[4-6],近几年在无线电通信中应用越来越广,有必要对它进行设计和它对系统性能的提升做定量分析。

2 LDPC 理论研究分析

2.1 LDPC 校验矩阵

LDPC (n, k) 属于线性分组码,其中 n 表示码字序列比特长度, k 表示信源序列比特长度。校验矩阵 H 可以用来唯一地确定一个线性分组码,对于维数为 $m \times n$ 的校验矩阵来说,其矩阵中的每一列对应码字的对应位、每一行对应一个相应的校验方程。设码字 $c = [c_1, c_2, \dots, c_n]$ 则 c 和 H 满足校验码字方程式(2-1)

$$H \times c^T = 0 \quad (2-1)$$

本文的基校验矩阵采用 IEEE802.16e 标准^[7-8]中所采用准双对角线子矩阵结构设计的校验矩阵,如图 2-1 所示:

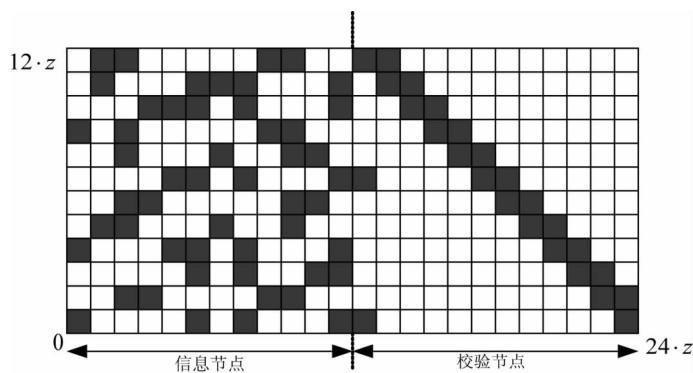


图 2-1 IEEE802.16e 标准中码率为 1/2 的校验基阵结构图

基校验矩阵 H_b 可以通过把自身每一个元素替换为一个由该元素值所决定的大小为 $z \times z$ 的子矩阵来得到最终的校验矩阵 H , 其中 z 为拓展参数, 具体的拓展关系如下表 2-1 所示

表 2-1 校验矩阵和基校验矩阵的对应关系

基校验矩阵元素值	校验矩阵子矩阵形式
-1	零方阵
0	单位矩阵
正数	单位矩阵循环右移阵

2.2 LDPC 编码算法

首先以拓展因子 z 的大小为长度间隔对信源序列 s 和校验序列 p 进行分段, 则信源序列和校验码序列可以写成

$$s = [s_1, s_2 \dots s_{k_b}] \quad (2-2)$$

$$p = [p_1, p_2 \dots p_{m_b}] \quad (2-3)$$

其中

$$s_i = [s((i-1)z+1), s((i-1)z+2), \dots, s(iz)], i=1, 2, \dots, k_b \quad (2-4)$$

$$p_i = [p((i-1)z+1), p((i-1)z+2), \dots, p(iz)], i=1, 2, \dots, m_b \quad (2-5)$$

LDPC 码字为一个长度为 n 的行向量 $c = [s, p]$ 。根据码字和校验矩阵的关系 $Hc^T = 0$, 可得

$$\begin{bmatrix} Z_{q(1,1)} & Z_{q(1,2)} & \cdots & Z_{q(1,k_b+1)} & \cdots & Z_{q(1,n_b)} \\ Z_{q(2,1)} & Z_{q(2,2)} & \cdots & Z_{q(2,k_b+1)} & \cdots & Z_{q(2,n_b)} \\ \vdots & \vdots & & \vdots & & \vdots \\ Z_{q(r,1)} & Z_{q(r,2)} & \ddots & Z_{q(r,k_b+1)} & \ddots & Z_{q(r,n_b)} \\ \vdots & \vdots & & \vdots & & \vdots \\ Z_{q(m_b,1)} & Z_{q(m_b,2)} & \cdots & Z_{q(m_b,k_b+1)} & \cdots & Z_{q(m_b,n_b)} \end{bmatrix} \cdot \begin{bmatrix} s^T \\ p^T \end{bmatrix} = 0 \quad (2-6)$$

$Z_{q(i,j)}$ 为基校验矩阵中相应元素 (i,j) 按照上述介绍的拓展方法展开的子矩阵。把公式(2-6)运算展开可以得到 m_b 个方程组,将所有方程组相加,可得

$$p_1^T = (Z_{q(1,k_b+1)} + Z_{q(r,k_b+1)} + Z_{q(m_b,k_b+1)}) \cdot \sum_{i=1}^{m_b} \sum_{j=1}^{k_b} Z_{q(i,j)} \cdot s_j^T \quad (2-7)$$

由于基校验矩阵的第 k_b+1 列的 $Z_{q(1,k_b+1)}$ 和 $Z_{q(m_b,k_b+1)}$ 相等,且 $Z_{q(r,k_b+1)}$ 为单位矩阵,上式可以化简为:

$$p_1^T = \sum_{i=1}^{m_b} \sum_{j=1}^{k_b} Z_{q(i,j)} \cdot s_j^T \quad (2-8)$$

将公式(2-8)代入下式可以得到

$$p_2^T = \sum_{j=1}^{k_b} Z_{q(1,j)} \cdot s_j^T + Z_{q(1,k_b+1)} \cdot p_1^T \quad (2-9)$$

根据已经求得的 p_1^T 、 p_2^T ,利用公式(2-6)可以依次计算出校验序列其他分段的 p_i 值

$$p_i^T = p_{i-1}^T + \sum_{j=1}^{k_b} Z_{q(i-1,j)} \cdot s_j^T, i=3,4,\dots,r,r+2,r+3,\dots,m_b \quad (2-10)$$

$$p_{r+1}^T = p_r^T + \sum_{j=1}^{k_b} Z_{q(r,j)} \cdot s_j^T + p_1^T \quad (2-11)$$

如果定义基校验矩阵的第 i 行元素的拓展子矩阵与消息序列相乘为:

$$\lambda_i = \sum_{j=1}^{k_b} Z_{q(i,j)} \cdot s_j^T \quad (2-12)$$

则校验序列 p 的计算可以整理如下:

$$\begin{cases} p_1^T = \sum_{i=1}^{m_b} \\ p_2^T = \lambda_1 + Z_{q(1,k_b+1)} \cdot p_1^T \\ p_{r+1}^T = p_r^T + \lambda_r + p_1^T \\ p_i^T = p_{i-1}^T + \lambda_{i-1}, i=3,4,\dots,r,r+2,r+3,\dots,m_b \end{cases} \quad (2-12)$$

2.3 LDPC 译码算法

假设发送的二进制序列为 $c = [c_1, c_2, \dots, c_n]$,接收端接收的信号是 $y = [y_1, y_2, \dots, y_n]$ 。 y 序列经过译码后得到二进制序列 \hat{c} 。下面是一组与译码算法相关的变量声明:

i — 表示第 i 个变量节点,对应于校验矩阵的第 i 列。

j — 表示第 j 个校验节点,对应于校验矩阵的第 j 行。

R_j — 表示与第 j 个校验节点相连的所有变量节点的集合,对应于校验矩阵第 j 行中元素非零位置的集合。

$R_{j/i}$ — 表示 R_j 中除去第 i 个变量节点其它变量节点形成的集合。对应于校验方程中除去第 i 个比特后第 j 行其它元素非零的位置集合。

C_i — 表示与第 i 个变量节点相连的所有的校验节点的集合。对应于校验矩阵第 i 列所有非零元素所在行的校验方程集合。

$C_{i/j}$ — 表示 C_i 中除去第 j 校验节点后其他校验节点形成程集合。对应于校验矩阵中除第 j 行以外第 i 列其它非零元素的位置所在行的校验方程集合。

$r_{j/i}(b)$ — 表示码字第 i 个比特为 b (b 的取值为 0,1)的情况下,码字的其它比特满足第 j 个校验方程的条件概率,这里称作校验节点信息。

$q_{i/j}(b)$ — 表示码字第 i 个比特为 b (b 的取值为 0,1)的情况下,该比特满足除第 j 行以外所有校验

方程的条件概率,这里称作变量节点信息。

$P_i(b)$ —表示码字中第 i 个比特仅由自身决定的其值为 b 的概率, b 的取值为 0,1。

本文采用的偏置最小和(Offset Min-Sum,OMS)算法的译码的基本步骤如下:

第一步 初始化:

$$v_{ij} = v_i^0 \quad (2-13)$$

第二步 校验节点更新:

$$c_{ij} = \prod_{j \in R_{ji}} \text{sgn}(v_{ij'}) \cdot \max(\min(|v_{ij'}| - \beta), 0) \quad (2-14)$$

第三步 变量节点更新:

$$v_{ij} = v_i^0 + \sum_{j \in C_i/j} c_{ij} \quad (2-15)$$

第四步 似后验概率更新:

$$v_i = v_i^0 + \sum_{j \in C_i} c_{ij} \quad (2-16)$$

第五步 比特判决:

如果满足 $v_i > 0$,则译码结果 $\hat{c}(i) = 1$;如果不满足则判定 $\hat{c}(i) = 0$ 。然后判断码字和检验矩阵是否满足 $H\hat{c}^T = 0$,如果满足则认为译码成功,译码过程结束同时输出译码结果;若不成立且此时迭代次数还没有达到规定的上限则返回到步骤 2 继续迭代,若译码不满足校验矩阵关系且此时迭代次数达到上限,则译码结束,输出错误码字。

其中 OMS 算法的初始化消息概率

$$v_i^0 = \ln\left(\frac{P_i(0)}{P_i(1)}\right) = \frac{2y_i}{\delta^2} \quad (2-17)$$

由于 OMS 算法采用一个偏移因子修正校验节点信息公式。图 2-2 展示了偏移因子 β 的变化对 LDPC 译码误码率的影响。

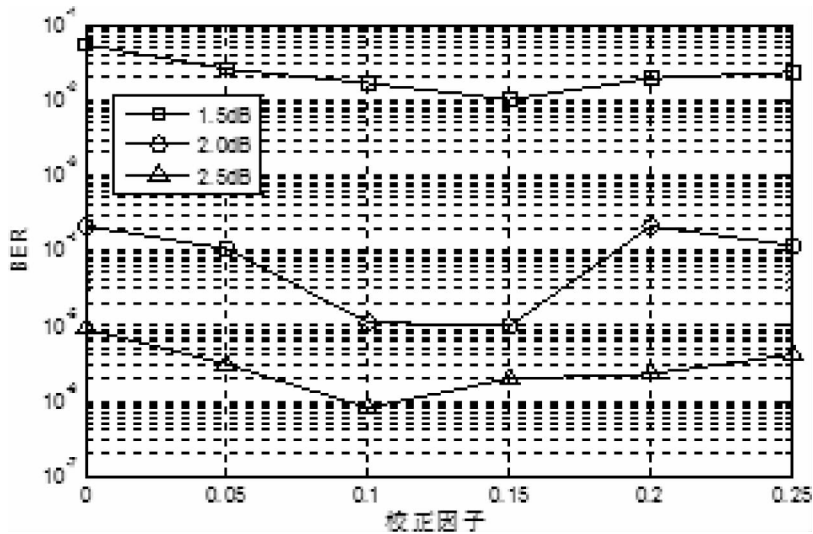


图 2-2 OMS 算法的偏移因子 β 对译码性能的影响曲线图

由上图可知,误码率在 0.1~0.15 范围内最低,考虑到硬件实现的方便性,本文 OMS 算法的偏移因子取值为 0.125。

图 2—3 给出了本文使用的 OMS 算法和大论文中其他算法性能仿真对比。由于 LLR BP 算法对 BP 算法的改进并没有对译码性能产生影响,因此两者的译码性能是相同,为了降低仿真运行的时间,本文仿真选取了运算相对简单的 LLR BP 算法。仿真中所有译码算法的最大迭代次数均为 14,码率均为 1/2 的 LDPC(960,480)进行仿真,分析误码率随信噪比变化,仿真结果处理后得到图 2—8。

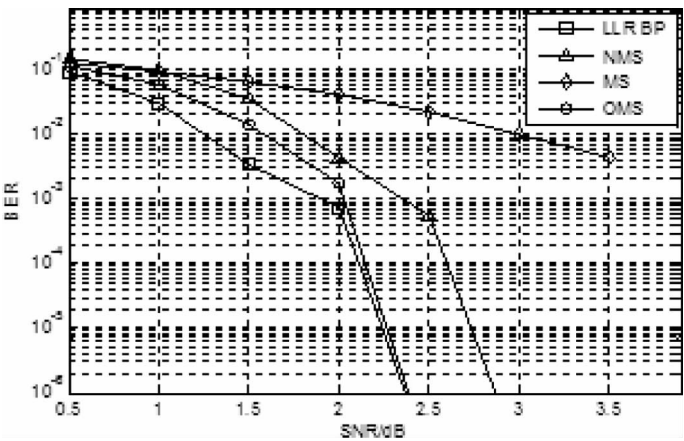


图 2—3 本文各种译码算法的性能对比图

分析上图可知,论文所讨论的各种译码算法除 MS 算法外,均具有良好的瀑布效应,其中 LLR BP 算法的瀑布效应最明显显示了优异的译码性能。在低信噪比的条件下,OMS 算法的译码性能较之 LLR 算法有一定的差距,但是随着信噪比的增大 OMS 算法的性能迅速提升有超过 LLR 算法性能的趋势。MS 算法的误码率性能随信噪比变换较为缓慢,瀑布效应不明显,在所有译码算中的性能最差,这与其化简的误差有关。考虑到实际系统设计的复杂度以及成本,论文选取 OMS 算法作为最终硬件实现的方案。

3 LDPC 编解码器设计与实现

3.1 编码器设计与实现

编码器的整体结构如图 3—1 所示,主要由三个模块组成:输入模块、编码模块和输出模块。输入模块负责和前级进行数据交互并为编码模块提供待译码消息序列;编码模块主要完成消息序列对应的校验序列计算^[8-10]。



图 3—1 LDPC 解码器整体结构示意图

输入输出模块对外接口采用握手方式进行数据交互,握手的时序图如 3—2 所示,其中 data 表示输入或输出的数据;data_valid 表示发端数据是否有效,高电平表示数据有效,低电平表示数据无效;data_ready 表示收端是否准备好接收数据,高电平表示准备好接收数据,低电平表示尚未准备好接收

数据。只有 $data_valid$ 和 $data_ready$ 同时为高电平,一次数据交互才能完成。输入模块和输出模块分别对应接收端和发送端。

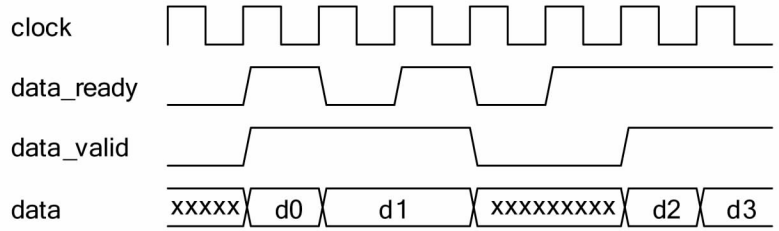


图 3-2 输入输出接口握手信号时序图

编码器设计采用拓展因子 $z=40$ 、码率为 $1/2$ 的准循环结构校验矩阵 H 。编码算法采用快速迭代译码算法。为了方便硬件设计的实现,对公式(2-8)校验信息序列 p_1 的计算修正如下

$$p_1^T = \sum_{j=1}^{12} \sum_{i=1}^{12} Z_{q(i,j)} \cdot s_j^T \quad (3-1)$$

在硬件实现过程中,当第 j 组的 z 比特消息序列 s_j^T 进入编码器,先计算 s_j^T 与基校验矩阵 H 第 j 列第 i 行元素所对应的拓展子矩阵 $Z_{q(i,j)}$, ($i=1,2,\dots,12$) 的乘积

$$p(i,j) = Z_{q(i,j)} \cdot S^T j \quad (3-2)$$

该乘积运算由下文所述的矩阵乘法模块完成,然后计算第 j 列所有行对应的 $p(i,j)$, $i=1,2,\dots,12$ 的累加值

$$p(j) = \sum_{i=1}^{12} p(i,j) \quad (3-3)$$

当 $j=1$ 的时候,即消息序列为第 1 组时候,把 $p_s(1)$ 的值锁存在寄存器中用以初始化寄存器。后续的 $p_s(j)$, $j \neq 1$ 与寄存器的值相加并继续存储在该寄存器中,此时寄存器的值为

$$S_j = \sum_{n=1}^j p_s(n) \quad (3-4)$$

当该消息序列最后一组进入编码器, S_{12} 就是校验序列 p_1 的值,利用另外一个寄存器锁存此值以便后续的校验序列计算。公式(3-3)中的 $p(i,j)$ 按照类似的操作可以计算出所有公式(2-11)所示的 λ_i 值。根据编码算法的公式(2-12),在时钟的驱动下依次计算其他校验序列分组。本文的设计采用如图 3-3 所示的流水线结构,当一组消息序列完全进入编码器后,计算其对应校验序列所需的参数都能够确定并锁存,在计算校验序列的同时下一组待编码消息序列可以继续进入编码器。因此流水线结构的编码器能够提升一倍的编码效率。

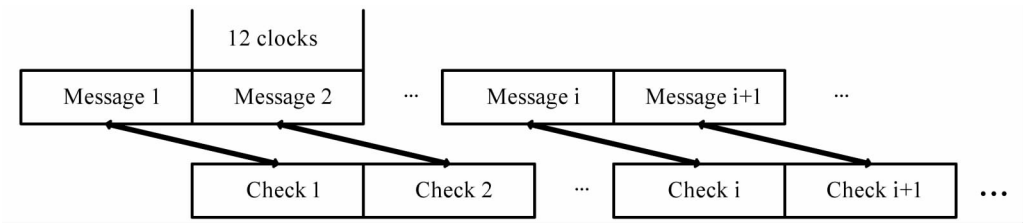


图 3-3 LDPC 编码器流水线设计示意图

Modelsim 仿真中使用的消息序列为 MATLAB 生成的随机序列,表 3-1 展示了测试中所使一个消息序列以及其对应校验序列分组。测试中消息序列的长度为 480 比特,为了方便与编码器的输出比较将消息序列和校验序列划分 80 比特一组。

表 3-1 仿真测试所使用码字的消息序列分组和其对应的校验序列分组

组号	消息序列(十六进制)	组号	校验序列(十六进制)
1	7BDD4B17994EBB93CA99	1	86428C9F7778190D5189
2	90A49C3570BCD5133403	2	099B5A10140D6109D207
3	2536E1144D99DC6E658E	3	62F6AB70ED12B3E827A8
4	CE22F7E08CFD701E9B1B	4	52E0D7022DC021145B2F
5	304514BE5E7F54FFD4ED	5	47399DCF9D5A32610AEA
6	A3FA1931D6239470F5CF	6	C4DB311728CFA42EAF79

Modelsim 仿真验证时,采用 MATLAB 产生的随机消息序列进行多次测试确保编码器工作的正确性。从图 3-4 可以看出,复位操作后,编码器第一组 80 比特的消息序列在第一个 40 比特消息序列进入后的第 5 个时钟后开始输出,说明编码器从输入到输出有 5 个时钟周期的延迟(该延迟是消息 FIFO 的输入到输出转换所需要时间)。第一个码字序列分组的输出为 0x7BDD4B17994EBB93CA99 这在上表中消息序列分组 0 一致,说明输入模块前端数据接口、输出模块中消息 FIFO 和控制模块中消息读取输出功能工作正常。

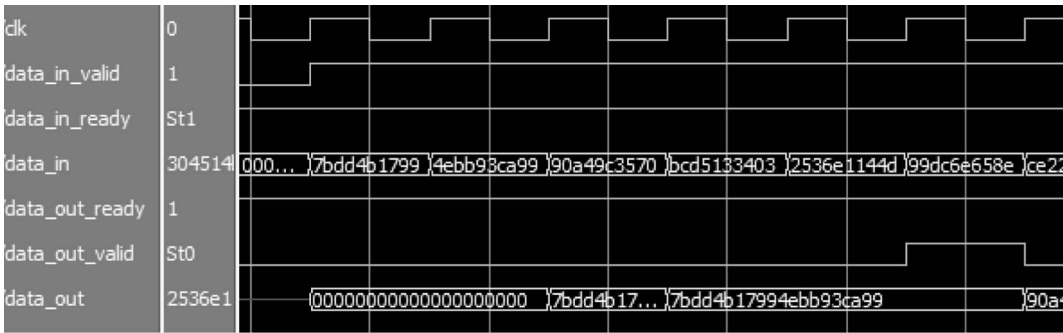


图 3-4 LDPC 编码器复位后, 码字的第一个消息分组(消息分组 0)输出仿真示意图

图 3-5 中标记线之前为当前码字序列中消息序列的最后一个分组,其后为该码字序列中的第一个校验序列。它们的值和上表中的消息分组 5 以及校验分组 0 一致,表明设计中的编码模块和输出模

块均能正常工作。

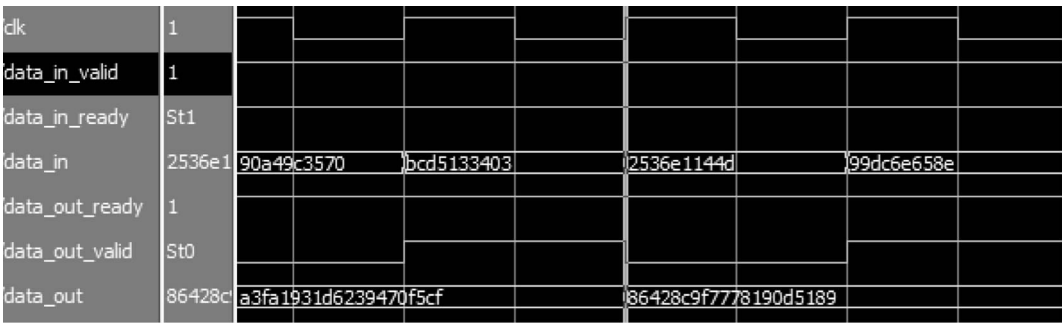


图 3—5 LDPC 编码器复位后,码字的第六个消息分组(消息分组 5)和第一个校验分组(校验分组 0)输出仿真示意图

图 3—6 为编码器满负荷工作时的仿真图,仿真中不停的往编码器输入消息序列,输出端的负载一直都处于准备好接收状态,标志线之前为上一码字的最后一组校验序列分组其值为 0xC4DB311728CFA42EAF79,标志线之后为下一个码字的第一组消息序列分组其值为 0x7BDD4B17994EBB93CA99,两者均和上表中的 MATLAB 仿真结果一致。输出信号 data_out_vali 一直为高电平表明编码器在按照流水线的设计意图连续的正常的工作,验证了设计的正确性。

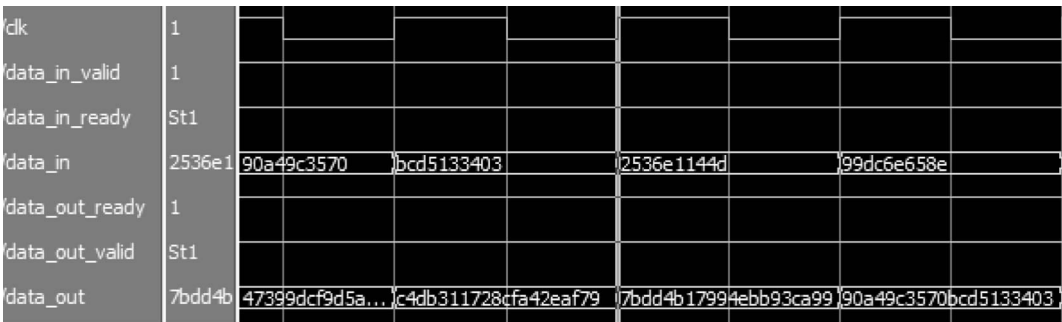


图 3—6 LDPC 编码器全速率工作后第 n 个码字的第六个消息分组(消息分组 5)和第一个校验分组(校验分组 0)输出仿真示意图。

代码仿真完毕后使用 Quartus II 11.0 进行编译综合,表 3—2 展示了本文编码器综合结果和其他文献综合结果的对比情况。

表 3—2 本文编码器和其他文献编码器设计对比一览表

编码器	文献 ^[11]	本文
Device	EP1S40B956C6	EP4CE115F29C7
Total registers	9090	5477
Total memory bits	11573	12384
吞吐率(系统时钟 50Mhz)	1Gpbs	2Gpbs

从表 3-2 可以看出,与文献^[11]相比,本文设计的编码器节约了 39.7%左右的逻辑资源,存储器的消耗要比文献中的设计略大 7%。但是在同样 50Mhz 工作时钟下,本文设计的吞吐率能够达到 2Gpbs,比文献高出一倍。本文编码器设计以较低的时钟实现了较高的吞吐率,在同样应用场景下,有利于时序的优化同时降低了整个系统的功耗。

3.2 译码器设计与实现

根据译码器的功能设计中把译码器划分为六个模块:输入缓存模块、控制器、信息存储模块、校验节点更新模块、变量节点更新模块和输出缓存模块。译码器接口设计采用和编码器一致的握手协议,译码器的整体结构如下图 3-7 所示。

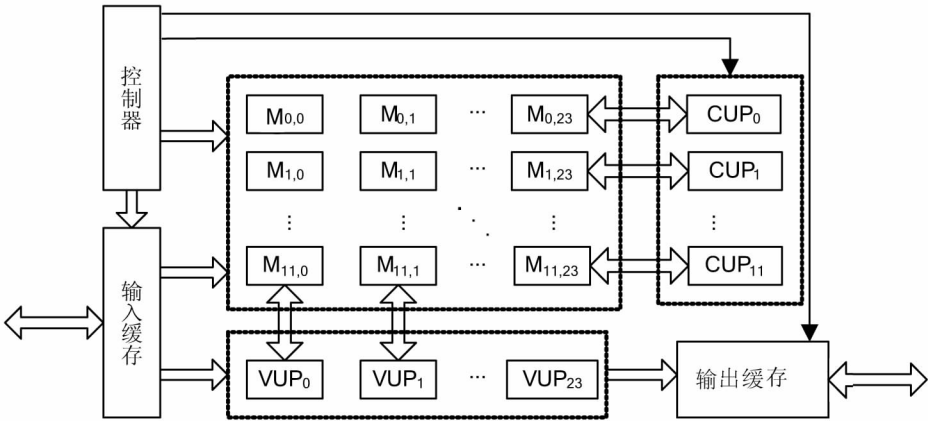


图 3-7 LDPC 译码器设计的整体结构示意图

FPGA 设计中节点信息和变量信息可以使用双口 RAM 交替存储^[12]。每一列的子矩阵对应一个变量节点信息更新(VUP)模块,每一行的子矩阵对应一个校验节点信息更新(CUP)模块。通过对读写地址的调度使得变量节点信息和校验节点信息在 RAM 中交替存储。由上述分析可知,设计中行列的并行度分别为 12 和 24 对应于校验基阵的大小。这种设计结构的硬件复杂度较低,数据处理效率高。

由于译码设计的复杂度较高,为了方便设计的仿真和调试,设计之前使用 MATLAB 搭建 C 语言模型。Modelsim 仿真过程中每个模块的功能和 C 语言模型进行严格比对确保一致性。最后在 FPGA 开发板完成在线测试。表 3-3 给出了硬件测试中所使用的一组码字,该码字引入错误后送入译码器进行译码。为了方便对比,输出的译码结果转换为 8 比特一组。引入错误时,把码字的前 48 比特全部置为 1。

表 3-3 译码器测试使用的正确码字和引入错误的码字列表

组号	正确码字	引入错误的码字
1	01110110	11111111
2	11011011	11111111
3	11001100	11111111
4	00010000	11111111

组号	正确码字	引入错误的码字
5	00000000	11111111
6	10100000	11111111
7

注:为了观测方便输出数据宽度转换为 8 比特

图 3—8 为上述错误码字进入译码器后的输出仿真结果,从图中可以直观地看出,经过译码器处理后,数据的前 48 比特所引入的错误全部都被纠正,设计的正确性得到了验证。



图 3—8 LDPC 译码器纠错后的数据输出

代码设计仿真完毕后使用 Quartus II 11.0 进行编译综合,并与文献中其他译码器性能进行比较,具体比较如下表 3—4 所示。

表 3—4 综合结果和性能分析对比表

	文献 ^[13]	本文译码器
Device	EP2SGX90FF1508C3	EP4CE115F29C7
Total registers	14,080	10,812
Total memory bits	187,968	127,350
吞吐量	34.2 Mbps	38.6Mbps

注:吞吐量计算的环境为,迭代次数 15 次,系统工作时钟 50Mhz,拓展因子 z=40。

从上表可以看出,本论文设计的译码器无论在硬件资源消耗还是在性能上都比文献中的设计要优。本文设计的译码寄存器消降低了 30%,存储器消耗降低了 47.5%,在相同工作时钟下吞吐量提升了 12.8%。

4 LDPC 的 UV 通信实验

4.1 UV 通信实验平台

为了更好突出纠错码在无线光通信系统中的作用,实验中采用 RS(18,10)和 LDPC(960,480)两种纠错码。实验采用控制变量法的方式进行。对于紫外光通信来说,接收端的信噪比容易受到通信距离和非视距的程度(仰角 α)的影响,如图 4—1 所示。

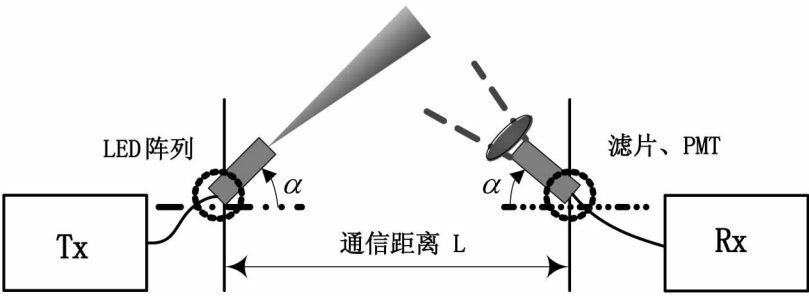


图 4-1 UV 通信非视距通信示意图

实验在不同仰角下,分析通信距离和系统误码率的关系。实验中的仰角分别取四种情况:0°对 0°、10°对 10°、20°对 20°、30°对 30°。0°对 0°为视距模式下的通信情况,其它三个角度对,反应了非视距的程度。考虑到实验执行过程中环境随机性和和人为误差,误码率的统计是基于大量实验的平均。表 5-1 详细的罗列了本文实验中系统的各项参数。

表 4-1 UV 通信系统参数表

参数	值
波长	265 nm
Tx 平均功率	1~10mW
PMT 灵敏度	62mA/W
调制方式	OOK
发送速率	500kbps
PMT 增益	4.9x10 ⁵
信道编码	LDPC,RS
采样率	8MHz

4.2 实验结果与分析

图 4-2 展示了视距(LOS)条件,采用编码和非编码的误码率和通信距离的关系曲线图。如果以可靠通信的误码率 10^{-6} 作为衡量标准,未采取编码的系统的可靠通信距离只能达到 22 米。采用 RS 编码的系统通信距离可以达到 27 米左右,通信距离提升大概 23%。采用 LDPC 编码的系统,从目前测试的距离范围内未发现任何误码,其可靠通信距离大于 50 米。

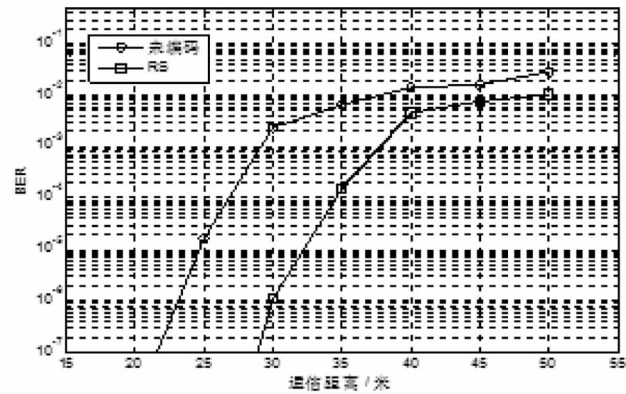


图 4-2 LOS 条件下编码前后误码率和通信距离关系图

在非视距模式下,收发仰角为 10° 对 10° 的情况下,未编码系统的通信距离只有 17 米左右,RS 编码提升系统通距离 8 米左右。采用 LDPC 编解码的系统的通信距离能够达到 36 米左右。

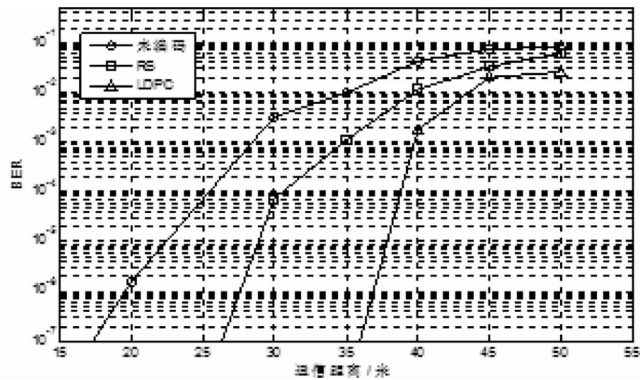


图 4-3 NLOS 条件下编码前后误码率和通信距离关系曲线图,收发仰角 10° 对 10°

在非视距模式下,收发仰角为 20° 对 20° 的情况下,未编码系统的通信距离缩减至 12 米左右,RS 编码的通信距离大约为 16。采用 LDPC 编解码的系统的通信距离能够达到 25 米。

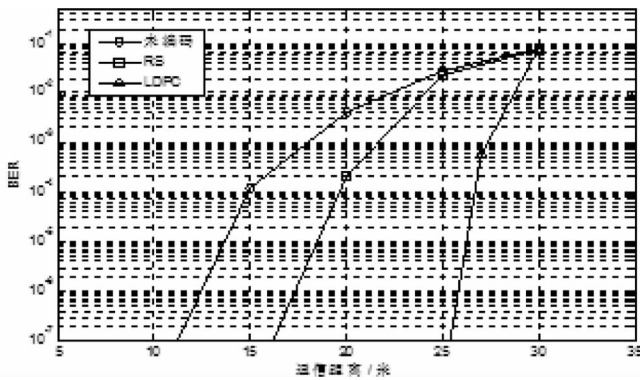


图 4-4 NLOS 条件下编码前后误码率和通信距离关系曲线图,收发仰角 20° 对 20°

在非视距模式下,收发仰角为 30°对 30°的情况下,未编码系统在所有通信距离下均存在很大的误码率,无法完成可靠通信。RS 编码的通信距离只有不到 6 米。采用 LDPC 编解码的系统的通信距离能够达到 10 米。

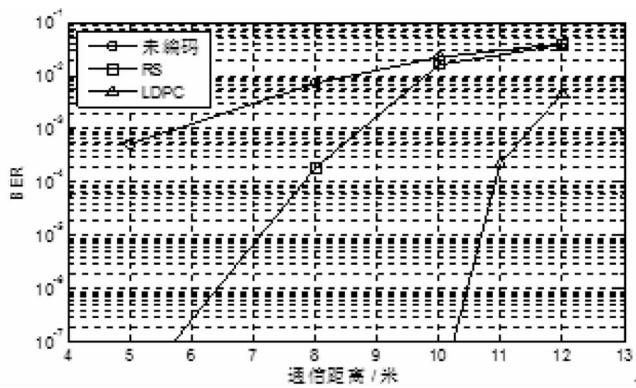


图 4—5 NLOS 条件下编码前后误码率和通信距离关系曲线图,收发仰角 30°对 30°

由表 4—2 可以看出,采用信道编码的 UV 通信系统的通信距离得到了极大的延长。在收发仰角为 10°对 10°的情况下,RS 信道编码提升了系统 42%的通信距离,LDPC 信道编码大约延长了 89%的通信距离。在仰角为 20°对 20°的情况下,RS 码增加通信距离与仰角为 10°对 10°的情况下相同,但是 LDPC 却能够增加通信距离 116.6%。并且随着信道环境的越来越恶劣,LDPC 编码器纠错性能越来越得到体现。

表 4—2 UV 通信实验中,不同仰角下信道编码前后的可靠通信距离表

仰角	未编码	RS	LDPC
0°	22 米	30 米	>50 米
10°	19 米	27 米	36 米
20°	12 米	17 米	26 米
30°	—	6 米	10 米

注:本文可靠通信定义为误码率小于 10⁻⁶

5 结论

通过把信道编码引入 UV 通信系统,系统的误码率情况能够得到有效的改善,同时增加了可靠通信的距离。实验和仿真表明:RS 码能够增加 42%的通信距离,LPDC 码大约能够增加一倍的通信距离。不过值的注意的是采用 LDPC 需要消耗大量的硬件资源,这可能阻碍 LDPC 码在紫光通信的实际应用。但是随着现场可编程逻辑阵列(FPGA)的快速发展和低廉化,这种阻碍将会越来越小。

参 考 文 献

- [1] 陈薛刚. 紫外光通信的军事应用研究. 国防科技, 2010, 3: 24~27
- [2] 陶立强, 梁菁等. 紫外光通信系统光源技术发展研究. 电子科技, 2011, 8: 124~127
- [3] 苟加志. 日盲紫外光通信大气传输特性研究: [硕士学位论文]. 重庆: 重庆大学, 2009
- [4] R. G. Gallager. Low Density Parity Check Codes. Cambridge, MA: M. I. T. Press, 1963
- [5] D. J. C. Mackay, R. M. Neal. Near Shannon Limit Performance of Low Density Parity Check Codes. Electronics Letters, 1997, 33(6): 455~457
- [6] D. J. C. Mackay. Good Error-correcting Codes Based on Very Sparse Matrices. IEEE Trans. Inform. Theory, 1999, 45(2): 399~431
- [7] IEEE 802.16e-2005 and IEEE Std 802.16-2004/1-2005 (Amendment and Corrigendum to IEEE Std 802.16-2004), IEEE Std 802.16e-2005, 2006
- [8] M. E. C. Fossorier. Quasi-cyclic Low-density Parity-check Codes from Circulant Permutation Matrices. IEEE Trans. Inform. Theory, 2004, 50(8): 1788~1793
- [9] T. J. Richardson, L. Urbanke. Efficient Encoding of Low-density Parity-check Codes. IEEE Trans. Inform. Theory, 2001, 47(2): 638~656
- [10] 肖扬. Turbo 与 LDPC 编解码及其应用. 北京: 人民邮电出版社, 2010, 106~128
- [11] 甘振华. IEEE802.16e 标准 LDPC 码编译码器设计与 FPGA 实现: [硕士学位论文]. 上海: 上海师范大学, 2012
- [12] 张函隽. 基于 WiMax802.16e 的 LDPC 码解码器的设计与优化: [硕士学位论文]. 上海: 上海交通大学, 2008
- [13] 杨建平. IEEE802.16e 标准 LDPC 译码器 FPGA 设计和实现: [硕士学位论文]. 成都: 西南交通大学, 2010

The Research and Implementaion of LDPC Codec in the Optical Communication

Zhang Feng

(School of Electronic and Information Engineering, North China Univ. of Tech., Beijing, 100144)

Abstract Noise has an obvious interference on the signal in Ultraviolet (UV) communication system. For purpose of the improvement of communication quality, the paper implements an improved Low-Density Parity-Check (LDPC) codec algorithm and simulates the performance of non-line-of-sight (NLOS) in UV channel model. An experiment is conducted on LDPC(960,480) and RS (18,10) test bed. The result shows that under the given power and 10^{-6} bit error rate (BER) communication distance increases 42% with RS, 100% with LPDC improving the communication performance effectively.

Key words UV communication; LDPC; RS; NLOS; BER

基于 Zigbee 的病人呼叫系统的设计

研究生 董婉青

导师 刘 红

(北方工业大学电子信息工程学院 北京 100144)

摘 要 为了便于医护人员实时追踪病人位置信息,系统结合 Zigbee 和 Web 技术优势,采用软、硬结合的开发方法,设计了一种实用的便于医生使用的无线医疗监测系统。系统硬件选用 CC2430/31 芯片,软件开发使用 IAR、VC++6.0、Dreamweaver 等环境编写代码实现 Zigbee 无线定位、串口数据处理、Web 访问等功能。经测试,系统可实现在指定区域追踪移动节点位置,并可通过 Web 访问方式查看移动节点实时位置信息。当移动节点离开指定监测区域时系统报警提示。本系统在国内监护病人方面实现了创新,为医护人员对病人的护理提供更有力的保障。

关 键 词 Zigbee 技术; Web 技术; 定位; 实时监测; 医护

作者简介 董婉青,女,1989 年出生,硕士研究生,主要研究方向为无线传感网络。

E-mail: dongwq0909@foxmail.com

1 绪论

随着人们对健康水平和生活质量要求的日益提高,人们对医疗护理的需求也逐步增多。我国作为人口大国,目前医院护理人员的数量远远不能满足大家的需求,所以利用科学技术提高我国医院的医疗护理水平迫在眉睫。对于大型综合性医院住院病人繁多,住院部将病人按照病种进行分类管理,不同病种的病人应避免互串病房,对这些病人在指定区域的位置监测显得尤为重要。传统的人工管理病人方式远远不能满足需求,为此本文提出了基于 Zigbee 的病人呼叫系统,可实现无线监测病人位置,并通过上位机和 Web 访问方式实时掌握病人位置,当病人离开指定监测区域时,系统报警提示,医护人员可以及时采取行动。系统具有移动性、扩展性和灵活性,节省了传统工作模式中浪费大量人力物力的工作方式,为病人提供更人性化、智能化的护理服务,同时系统在医疗领域上实现了应用创新。

2 系统相关技术

系统采用 Zigbee 和 Web 等相关技术进行系统方案设计以及具体功能的实现,下面将对这两种技术进行介绍并阐述 Zigbee 技术与 Web 技术相结合所具有的优势。

2.1 Zigbee 技术

Zigbee 技术以其低复杂度、低功耗、低数据速率、低成本、自组织等特点而闻名。Zigbee 协议栈在 PHY 层和 MAC 层规范之上实现了 NWK 层和 APL 层,并在 APL 层内提供了 APS 和 ZDO。AF 中具有用户自定义的应用对象,本系统的定位功能就是通过在应用层编写相应代码实现系统功能的。

系统开发使用 TI 公司配合 Zigbee 硬件开发而推出的 Z-stack 协议栈,该协议栈提供了 OSAL (操作系统抽象层)的协议栈调度程序,通过调用 API 接口进行开发。OSAL 程序由 main 函数执行系

统初始化,对协议栈中每一层的事件处理函数执行轮转查询式的操作,整个协议栈会按照时间顺序有条不紊的运行。

2.2 Web 技术

随着互联网络技术的发展,Web 技术得以产生并迅速发展。Web 技术具有图形化界面、动态页面、分布式结构和兼容性强等优势,适用于本系统定位信息查看功能的使用。基于 Web 的网络架构分为 C/S 和 B/S,与 C/S 架构相比,B/S 架构可以为使用者提供不限时间、地点、接入方式等方面限制的方式实现对数据库的操作,因此系统选择 B/S 架构。

2.2.1 动态网页技术

动态网页技术能通过后台与用户交互,完成用户查询、提交等动作。动态网页技术具有很强的交互性,网页可以动态地根据用户请求来进行相应的响应与改变。目前主流的动态网页技术有 JSP、ASP、PHP^[1]等,综合三种主流动态网页技术的优劣势和网站需求等情况,本系统选择 PHP 技术作为动态网页开发的脚本语言。

2.2.2 Apache+PHP+MySQL 技术

Apache 服务器兼容性强、操作简单、运行速度快而且比较稳定,所以系统选择 Apache 作为 Web 服务器。PHP 技术具有很强的交互性,可以动态地根据用户请求来进行相应的响应与改变,所以选择 PHP 脚本语言。MySQL 体积小、速度快、安全性强、成本低而且易使用,所以选择 MySQL 数据库。基于三种技术的优势,系统选择在 windows 下搭建 Apache+PHP+MySQL 环境作为 Web 网站的开发环境。

3 系统工作原理

系统对携带移动节点的病人在指定区域进行定位^[2],网关节点接收移动节点坐标信息传送至上位机,上位机软件对数据处理并存储至数据库。Web 网站使用数据库中定位信息相关数据实现移动节点位置显示功能。系统主要通过 Zigbee 无线通信功能实现移动节点定位,下面将主要对定位部分工作原理进行阐述。

3.1 系统功能

本系统可为医生、病人提供集定位、查询等智能化护理服务为一体的平台,系统具体功能有以下三点:

- (1)病人随身携带移动节点,系统可在指定区域内对病人位置进行实时追踪;
- (2)登录 Web 查看病人实时位置信息,当病人离开监控区域系统报警提示;
- (3)可通过网站对使用者及病人信息进行管理,包括权限管理、数据查看、病人进出监控区域记录等。

3.2 系统组成

根据系统的功能需求采用软、硬件相结合的开发方法对系统进行设计,按照系统功能可将系统分为定位、数据处理、信息显示三部分,系统组成图如图 3.1 所示。

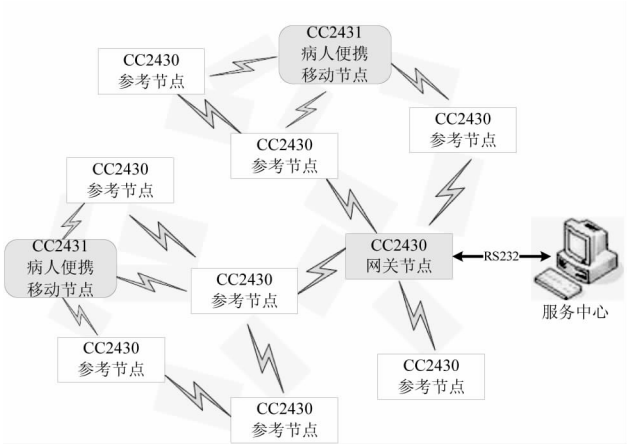


图 3.1 系统组成示意图

3.3 系统定位原理

系统主要实现病人在指定区域内的位置追踪功能,系统选用以 CC2430/31 芯片为核心的开发板,根据功能的不同可配置成移动节点、参考节点、网关节点,各节点间相互配合完成定位。现有的室内定位技术主要有近似法、几何法和 RSSI 法等。

RSSI 是通过接收信号强度所具有的变化规律来测定发送端和接收端距离的方法来实现定位功能的。因为系统选用的 CC2430/31 芯片可提供 RSSI 值,所以本系统使用基于 RSSI 测距技术来实现定位功能。RSSI 与距离 d 关系如式(3.1)所示:

$$RSSI = -(10 \times n \times \lg d + A) \tag{3.1}$$

式中, n 为传播系数(常数); d 表示接收端与发送端的距离; A 表示与发送者相距 1 m 时信号的强度值。

定位的精确程度取决于参数 n 与 A 的取值。参数 A 与实际环境有关,通过实际测量与发送者相距 1 米处的 RSSI 值得到 A 值。参数 n 也与实际环境相关,它是描述随着距离增加而递减的信号强度的参量,可在实际环境中通过多次测量取得最适合环境的 n 值。CC2431 所具备的硬件定位引擎就是根据这些参数以及参考节点坐标值即可计算出自身所在位置坐标值。

4 系统硬件设计

系统硬件部分主要完成定位、数据传输、数据处理等功能。按照硬件实现功能,可将系统节点分成移动节点、参考节点和网关节点。系统硬件组成图如图 4.1 所示,其中移动节点使用的是以 CC2431^[6] 无线射频芯片为核心的开发板,参考节点和网关节点使用的是以 CC2430 无线射频芯片为核心的开发板,其中网关节点为方便观察状态,添加外设 LCD 液晶屏,用以实现显示功能。图中虚线表示两者通过无线连接方式,网关节点与 PC 机是通过 RS232 接口有线连接。

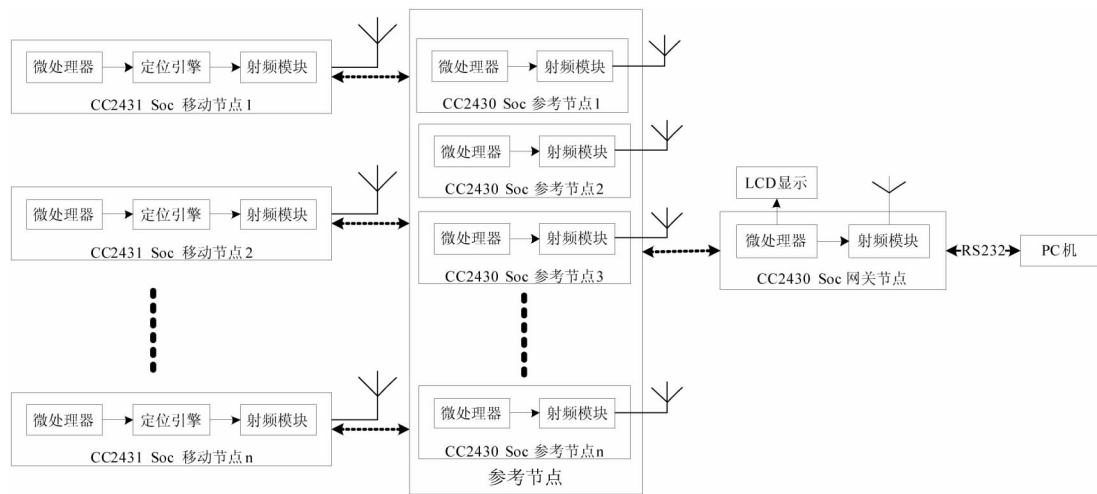


图 4.1 系统硬件组成图

4.1 参考节点

系统中有规律地布置在定位空间范围的为移动节点提供参考坐标值的这类节点就是参考节点。参考节点主要由 Zigbee 无线射频模块和微处理器组成的 CC2430 芯片及外围设备构成。

4.2 网关节点

系统中负责启动网络、向各节点发送配置信息、接收移动节点定位信息并将信息上传至上位机的节点是网关节点。网关节点主要由集 Zigbee 无线射频模块和微处理器于一体的 CC2430 外加 LCD 液晶屏组成,网关节点与 PC 机通过 RS232 接口相连,用于向上位机传送数据。

4.3 移动节点

系统中用来实现计算定位信息并将信息通过无线通信方式发送至参考节点功能的节点就是移动节点。移动节点是由集微处理器、定位引擎、射频模块为一体的 CC2431 片上系统组成的。

5 系统软件设计

系统软件设计主要包括移动节点、参考节点、网关节点功能^[7]的实现以及上位机软件的设计,系统使用 zstack—1.4.3—1.2.1 协议栈,编写应用层相关代码实现定位功能,网关节点通过 RS232 串口接收并处理数据,使用基于 MFC 编程实现数据处理功能。

5.1 定位节点软件设计

定位部分功能主要由移动节点、参考节点和网关节点相互协作完成。参考节点是系统中已知自身位置的静态节点,这类节点固定坐标布置在被定位区域,它们将自身坐标等信息通过无线通信方式发送至其它节点。当有移动节点进入到参考节点所覆盖的监控范围内,移动节点将会选择接收与其距离最近的几个参考节点的数据包进行接收,以获取附近参考节点处坐标及相应的 RSSI 值并将其写入定位引擎,通过定位引擎^[8]计算得到的自身位置坐标值。移动节点将自身坐标信息发送至网关节点,由网关节点通过 RS232 与电脑上位机软件相连,从而使上位机软件获取位置^[9]信息。定位部分的软件设计就是实现移动节点、参考节点、网关节点功能所进行的设计。

5.1.1 移动节点

移动节点 CC2431 的软件设计主要对与定位引擎相关的寄存器进行控制来实现移动节点定位^[10]

功能。定位引擎所需的输入值为附近参考节点所提供的坐标值、RSSI 值和与环境相关的参数 A 值、n 值等,定位引擎通过计算,输出自己所在位置的坐标值。移动节点工作流程图如图 5.1 所示。

5.1.2 网关节点

网关节点在定位系统中的角色是网络协调器。网关节点负责启动整个网络,所以它首先接收由上位机系统提供的各参考节点和移动节点的配置数据信息,并将这些信息对应节点功能发送给相应节点。然后网关节点接收各类节点的反馈信息传送至上位机,完成子节点与上位机的消息传递功能。网关节点所完成的工作都是通过应用层程序^[11]完成的,网关节点的流程图如图 5.2 所示。

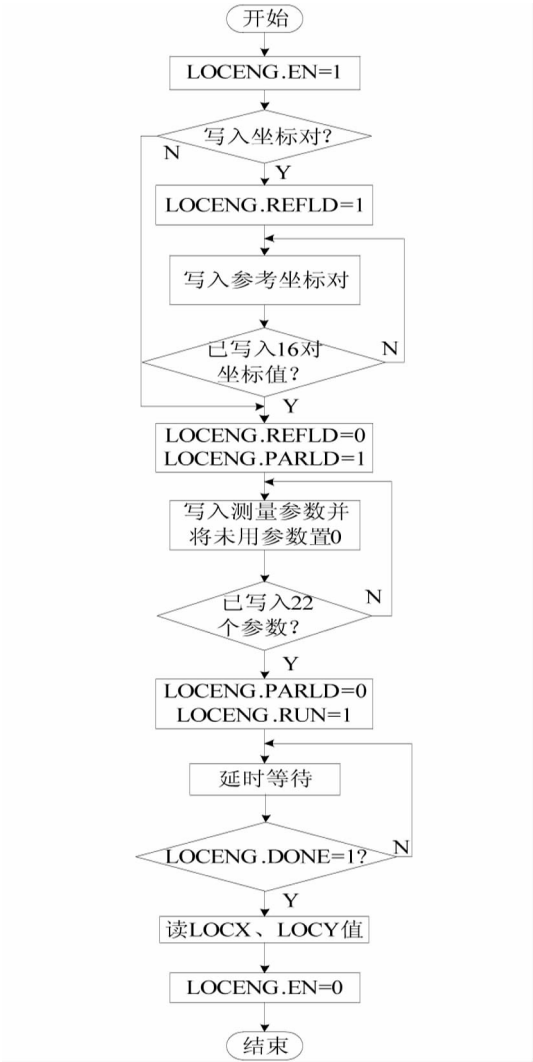


图 5.1 移动节点工作流程图

5.1.3 参考节点

将参考节点按照一定规律布置到被监控区域中,每个参考节点写有固定的坐标值,在系统中为进入监控区域的移动节点提供其自身的静态坐标和 RSSI 值。根据参考节点所完成工作对其进行软件设计,参考节点的工作流程图如图 5.3 所示。

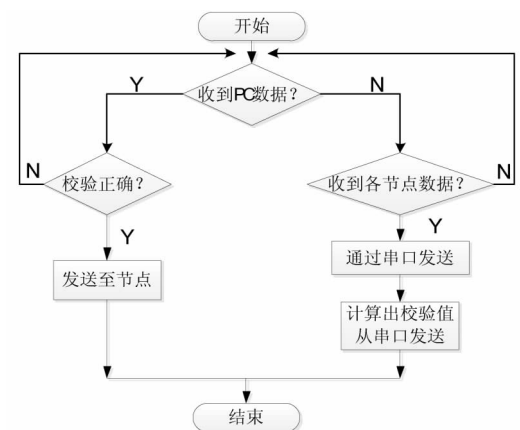


图 5.2 网关节点工作流程图

5.2 串口数据处理

网关节点通过无线通信方式接收移动节点坐标信息,并与计算机通过 RS232 接口有线相连,实现从串口读取数据并将信息存入数据库的功能^[12]。

系统使用串口通信类 CSerialPort 编写程序^[13],调用函数顺序为:调用 InitPort()和 StarMonitoring()打开串口,调用 StopMonitoring()和 ClosePort()关闭串口。串口通信通过轮询等待的方式等待数据接收或发送,从而实现对串口数据的处理。

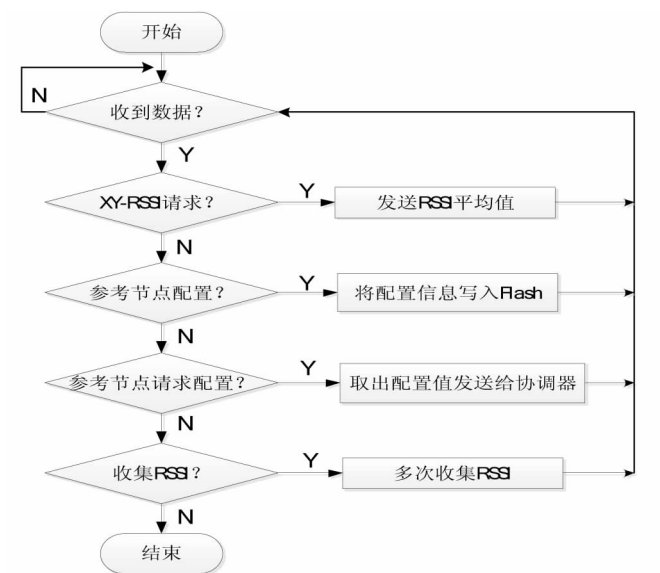


图 5.3 参考节点工作流程图

6 数据库设计

数据库是 Web 网站动态显示移动节点位置信息的基础,所以在设计 Web 显示^[26]移动病人位置信息时充分考虑对数据库的需求,对其功能进行分析,使用 MySQL 建立数据库,通过 phpMyAdmin 对数据库进行管理等相关操作,最终实现定位信息的存储、查询、使用等相关功能。

6.1 数据库功能需求

数据库主要完成对用户、病人、移动节点、参考节点等信息的添加、修改、查询操作,数据库需要具备的具体功能有管理员对用户、病人、移动节点具有操作权限;对新注册用户进行编辑、管理等操作;移动节点的信息存储、编辑等操作;参考节点相关信息存储、编辑等操作。

6.2 E-R 图

根据系统的功能需求,使用 E-R 图^[14]对数据库需求进行描述。数据库的实体为管理员、用户、病人、参考节点和移动节点,它们之间的关系有管理、编辑、设置、服务等。如图 6.1 为管理员与用户 E-R 图,图 6.2 为管理员与病人信息 E-R 图,图 6.3 为管理员与参考节点位置信息 E-R 图,图 6.4 为管理员与移动节点 E-R 图,图 6.5 为定位信息 E-R 图。

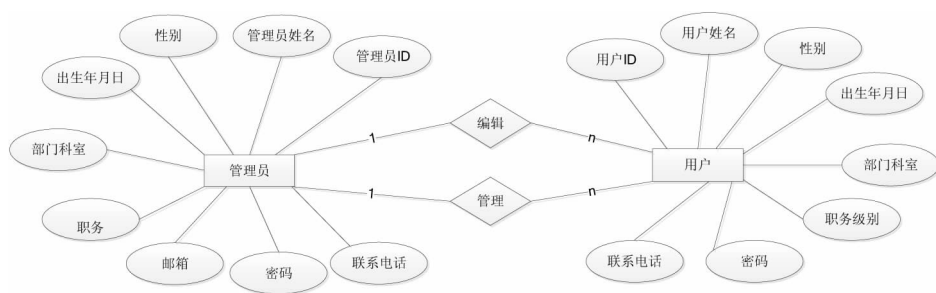


图 6.1 管理员与用户 E-R 图

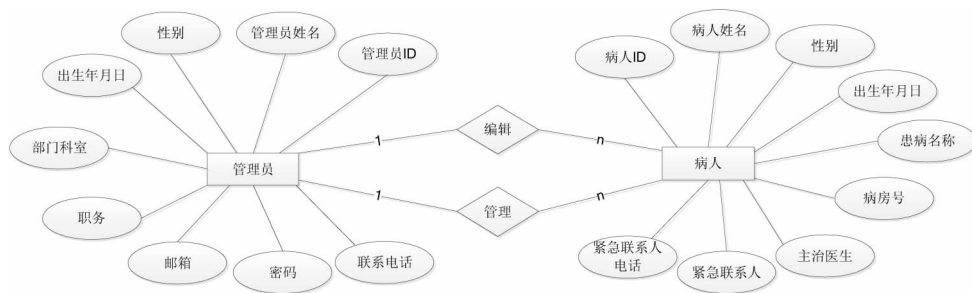


图 6.2 管理员与病人 E-R 图

6.3 数据字典

根据 E-R 图在 phpMyAdmin 环境下建立名为 apm 的数据库,包含以下数据字典:用户数据字典、病人数据字典、参考节点数据字典、移动节点数据字典、定位信息数据字典。

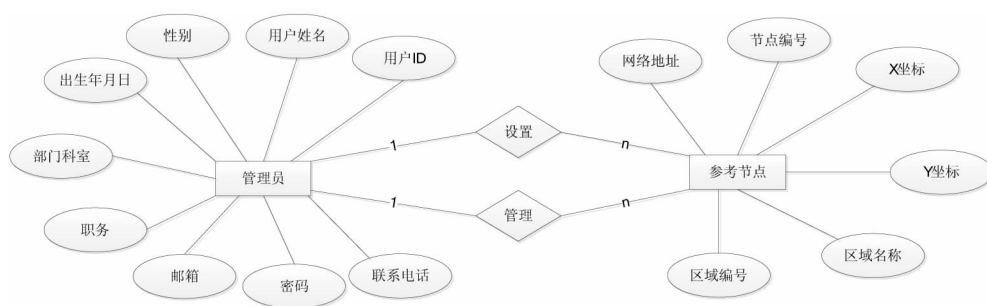


图 6.3 管理员与参考节点 E-R 图

(1)用户数据字典:用户数据字典用于记录用户信息,管理员利用字段名为“level”级别来区分使用权限。

(2)病人数据字典:病人数据字典用于记录病人个人信息,因为每位病人均手持移动节点所以“patientsinfo”表中字段名为“patientID”与“blindnodeinfo”表中用于标识移动节点 ID 名相同,以此来将移动节点与病人信息建立联系。

(3)参考节点数据字典:参考节点数据字典用于记录参考节点在被定位区域的分布情况及编号等信息。

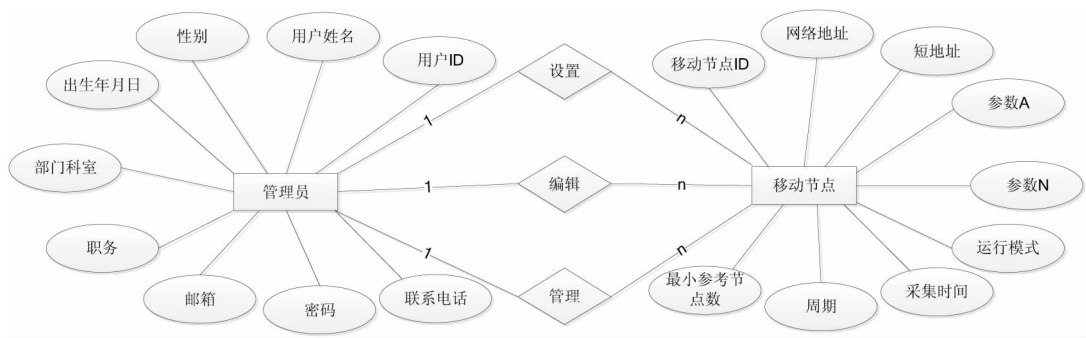


图 6.4 管理员与移动节点 E-R 图

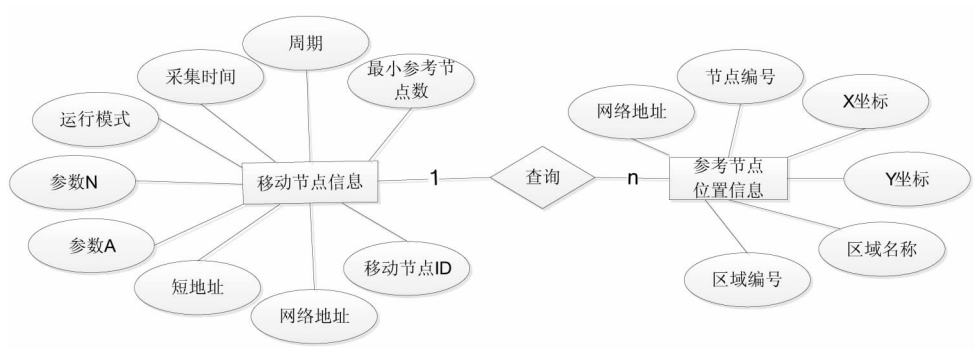


图 6.5 定位信息 E-R 图

(4)移动节点数据字典:移动节点数据字典用于记录病人携带的移动节点相关信息,运行模式的 0 值表示等待,1 表示自动发现;采集时间是发送请求后等待的时间;周期是自动扫描位置所用的时间;最少节点数是参与定位计算的最少参考节点的数量。

(5)定位信息数据字典:定位信息数据字典用于记录移动节点实时坐标,表名“patientID”与字典表“patientsinfo”中病人 ID 相同,用于把移动节点位置与病人相关联,以实现病人位置追踪功能。

7 Web 网站设计

Web 网站^[15]提供用户信息注册、登录、密码找回、浏览信息、查看定位信息等功能,医护人员可以通过登录网站查看病人实时^[16]位置信息,当病人离开其指定监测区域时,系统报警提示。

Web 网站以使用者实际需求为出发点,对网站结构进行设计。网站主页包含用户登录、注册、信

息浏览和搜索的功能模块,为实现以上功能所设计的网站结构如图 7.1 所示。

7.1 主页设计

网站主页主要实现用户登录、新用户注册、各种信息刊登以及相关信息搜索功能。用户登录与注册主要是用户与数据库交互的过程,系统使用 PHP 语言实现用户表单存储、使用等功能。当用户输入内容与数据库中数据表内容一致便可以成功登陆;新用户在正确提交注册信息后即可获得具有登录权限的用户名和密码。

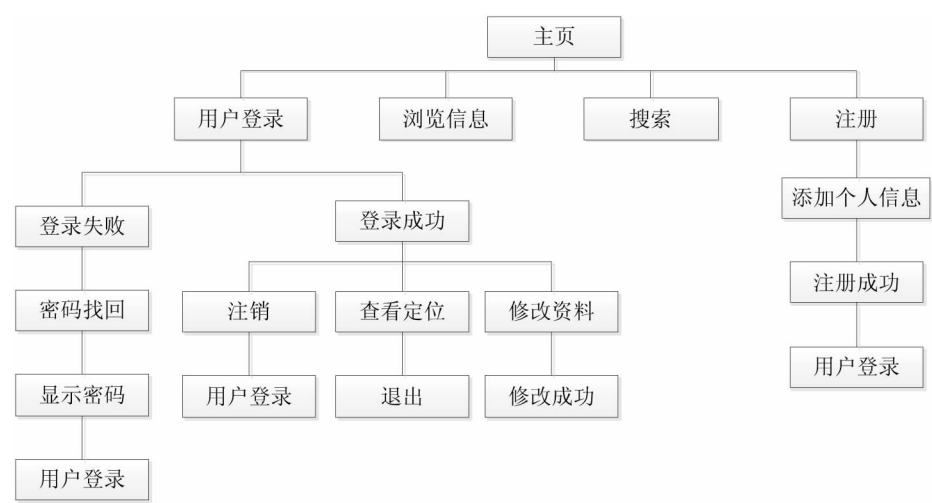


图 7.1 网站结构图

7.2 密码找回设计

考虑到用户使用网站的安全性,每位新用户注册时均设置了登录密码,但是用户在使用过程中不免出现忘记密码的时候,所以本网站设计了密码找回功能,为忘记密码的用户提供密码找回的服务。在用户注册时,按照相关提示存储了找回密码的提示问题以及相关答案,找回密码即根据用户提供的这些信息来确认是否为用户本人。

7.3 显示定位信息设计

Web 网页从数据中读取移动节点实时坐标信息,并通过定时刷新来显示移动节点实时的位置变化。使用 PHP 自带绘图功能,实现移动节点绘制功能。当移动坐标超过指定区域坐标范围,系统弹框报警提示。

8 系统测试

本系统模仿医院实际使用环境,使用本校三教 2603、2605 两间相邻实验室作为实验测试环境使用 1 个网关节点、8 个参考节点和 1 个移动节点进行黑盒测试。根据实际实验环境,以 2603 实验室为指定监测区域进行参考节点的布局,布局如图 8.1 所示。由于 2603 室内遮挡物较多,因此室内参考节点布置较密集,节点间以 3.2 米为间隔,布置在同一高度平面上。

上位机接收网关节点数据,经过串口处理程序将定位信息存储至数据库。经测试,接收数据符合协议标准,串口处理程序可实现预期功能。

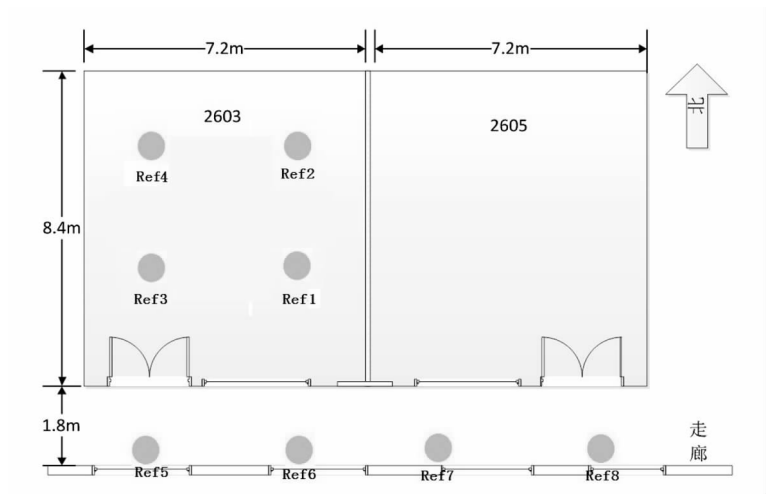


图 8.1 实验场景平面示意图

网站主页如图 8.2 所示,网站主页提供用户登录、注册、信息浏览、系统使用说明等功能。新用户可以通过单击“Register”按照网站提示要求填写好个人信息后即可获得登录权限。



图 8.2 网站主页

成功登陆后,根据网站页面提示查看移动节点位置情况,效果如图 8.3 所示,红点代表移动节点位置。其中 2603 为主监测区域,当病人离开主监测区域时如图 8.4 所示,系统报警提示。

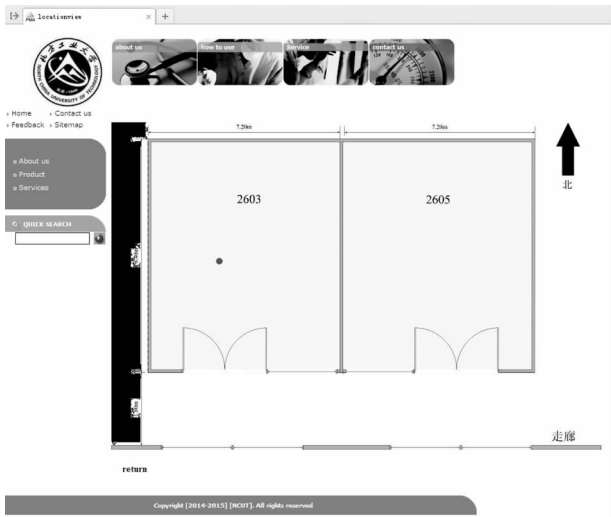


图 8.3 定位效果图

如果用户忘记注册密码,可以单击“Forgetpassword”输入用户名,正确输入注册时所填写的提示问题答案,即可找回登录密码。

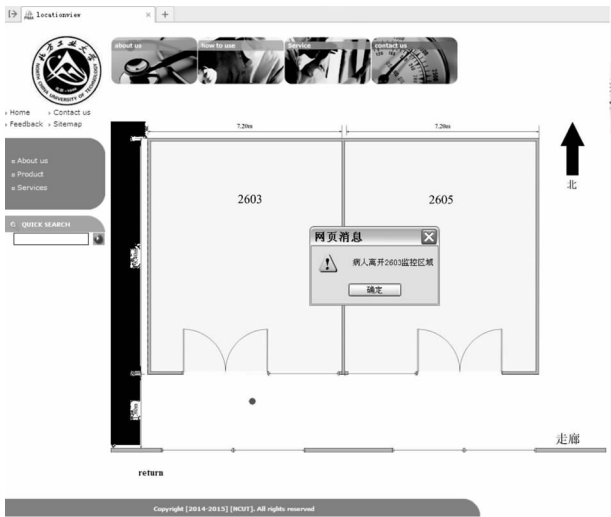


图 8.4 网页报警提示

通过黑盒测试,系统在实际应用环境中实现预期功能。使用者通过上位机和 Web 访问的方式对定位信息进行查询,大大方便了使用者在一定范围内对移动节点的有效监管。测试充分考虑病房区域环境情况,使用相连的两个实验室以及走廊作为实验区域,并将其中一间实验室作为指定监测区域,布置更多参考节点来提高定位精度,使其定位误差控制在 3—5 米。当病人离开指定监测区域时,系统报警提示。

9 结论

系统采用软、硬件结合的开发方法,设计了一种实用的便于医生实时监测病人位置的无线医疗监测系统,使用者可以通过上位机软件和登录 Web 网站两种方式查看病人位置信息,当病人离开指定监控区域时,系统报警提示。系统硬件部分使用以 CC2430/31 为核心的开发板,各节点主要通过 C 语言在 IAR 集成开发环境下使用 zstack-1.4.3-1.2.1 协议栈,编写应用层相关代码实现参考节点、移动节点和网关节点的功能。系统基于 MFC 编程实现串口数据处理并将定位坐标信息存储至 MySQL 数据库。Web 网站提供注册、登录、密码找回等功能,便于使用者不限时间地点查看病人位置信息。本系统在测试中取得了初步效果,但是由于硬件设备等条件的限制,仍需要进一步提高系统的定位精度。本系统提供集实时定位、显示、数据统计等功能为一体的位置监测系统,不仅可以用于住院病人,还可用于社区居家养老。与病人一样,老年人也属于高危人群,需要在社区指定活动区域实时掌握他们的位置情况,一旦离开活动区域可以第一时间发现并采取相关措施。综上所述,本系统在医疗护理方面实现了应用创新,为提升我国医疗护理水平方面提供了一定的参考。

参 考 文 献

- [1] 李谋平. 基于 PHP 的图书馆设备管理系统研究与开发[D]. 安徽:安徽大学硕士,2011
- [2] 曹琦. 基于 ZigBee 的社区居家养老系统的研究[D]. 陕西:西安工业大学硕士,2011
- [3] Texas Instruments CC2430 datasheet[EB/OL]. <http://www.ti.com>
- [4] Texas Instruments CC2431 datasheet[EB/OL]. <http://www.ti.com>
- [5] 陈晓娟. 医疗信息交互系统的设计与实现[D]. 湖北:武汉理工大学硕士,2012
- [6] AN042:cc2431 Location Engine[EB/OL]. <http://www.ti.com>
- [7] 黄新波,罗兵,刘存孝,黎小林,张伟,陶保震,李立涅. 采用 ZigBee 芯片的无线加速度传感器网络节点的实现[J]. 高电压技术,2010(8)
- [8] 梁芳. 基于 Zigbee 的监区人员定位研究[D]. 山西:中北大学硕士,2014
- [9] 李方. 基于 ZigBee 的位置指纹法室内定位技术研究[D]. 黑龙江:哈尔滨工业大学博士,2012
- [10] 杨文姬,王命延,苏武. 基于 RSS 场景指纹的 ZigBee 无线传感器网的定位方法[J]. 计算机与现代化,2010(10)
- [11] 杜文涛. 基于 WEB 的井下人员定位信息系统的研究[D]. 陕西:西安科技大学硕士,2009
- [12] 路娟. 基于 ZigBee 网络的电信机房动力环境远程监控系统的设计研究[D]. 山西:太原理工大学硕士,2012
- [13] 李明学,展影清,田由辉,张雅君. 基于 ZigBee 的井下人员定位系统的研究[J]. 机械制造与自动化,2013(3)
- [14] 徐良. 基于 ZigBee 的楼宇内群体体温监控系统的设计与实现[D]. 四川:西南交通大学硕士,2012
- [15] 蒋挺,赵成林. 紫蜂技术及其应用[M]. 北京:北京邮电大学出版社,2006. 238—251
- [16] 高明明,杨贵洲. 基于无线网络的矿井安全监控系统设计[N]. 辽宁工程技术大学学报(自然科学版),2011(6)

Design of Patient Call System Based on Zigbee

Dong Wanqing

(School of Electronic and Information Engineering, North China Univ. of Tech. , Beijing, 100144)

Abstract In order to enhance the medical staff in a real-time tracking patient location information, the system combined the advantage of ZigBee technology and web technology. The development method of the system uses the software and hardware to design a practical wireless medical monitoring system for medical staff. The CC2430/31 chip hardware were selected as the hardware, and the IAR, Dreamweaver, VC++6.0 were selected as the software development environment. Then write code to achieve the wireless positioning of Zigbee, serial data processing, Web access and other functions. After testing, the system can now specify the real-time location of the blind node and it can be accessed through Web. When the blind node leaves the designated monitoring area, the alarm system is prompted. The system has realized the innovation in the patient monitoring system in the country, it can provide more effective protection for patient care.

Key words zigbee technology; web technology; location; real time monitoring

10 位 10MHz SAR ADC 设计

研究生 史立韬

导师 戴 澜

(北方工业大学电子信息工程学院 北京 100144)

摘 要 逐次逼近型模数转换器(SAR ADC)以其低功耗和低面积开销的优点,在电子系统中得到了很广泛的应用。本论文设计了一种 10 位 10Ms/s SAR ADC;该电路内置的 DAC 使用全电容阵列设计,在电容阵列布局中采用新型算法来减少电容失配;通过电容的上极板采样信号,对电容阵列的开关逻辑电路进行优化,在一定程度上降低了功耗;比较器使用降低回馈噪声设计。最终对版图布局中各个部分中的寄生参数进行了优化,后仿真结果表明:使用 SMIC 0.18 工艺在 1.8V 电源下 SNDR 达到 59dB,即有效位达到 9.5 位,芯片面积为 0.6mm²。

关 键 词 逐次逼近模数转换器;比较器;电容阵列

作者简介 史立韬,男,1990 年出生,硕士研究生,主要研究方向为模数转换器。

E-mail: gl_jungly@foxmail.com

1 绪论

模数转换器是模拟与数字的接口,所以在现代电路中模数转换器处于很重要的位置。模数转换器的性能往往直接影响到整个电路的性能,与其他结构的模数转换器相比,逐次逼近寄存器型(SAR)模拟数字转换器(ADC)的结构是在中低采样速率和中等至高分辨率的应用中常用的结构。SAR ADC 具有一个显著的特点:功耗随采样速率而改变。这一特点与闪速 ADC 或流水线 ADC 不同,后者由于结构决定了其在不同的采样速率下具有固定的功耗。SAR ADC 这种可变功耗特性对于低功耗应用或者不需要连续采集数据的应用非常有利。例如:便携/电池供电仪表、笔输入量化器、工业控制和数据/信号采集等。SAR ADC 实质上是一种二进制搜索算法。因此,当内部电路运行在数兆赫兹(MHz)时,由于逐次逼近算法的缘故,ADC 采样速率仅是该数值的几分之一。

SAR ADC 的具体数字控制逻辑方法有很多,但是其核心思路都是二进制搜索算法,信号由采样保持电路,内置 DAC 输出信号与比较器进行比较。然后比较器对内置 DAC 的输出信号进行判断,根据比较器输出,控制内部 DAC 生成下一次输出结果,随后 SAR 控制逻辑移至下一位,进行下一次比较,寄存器从最高位 MSB 到最低位 LSB 依次改变。上述操作结束后 ADC 完成转换,N 位转换结果存储在寄存器内,然后随着下一次电路采样信号到来整体输出。

本文设计实现了一种内置 DAC 使用全电容阵列的 SAR ADC,采用动态比较器和电容上极板采样结构并使用单调电容开关,使控制逻辑结构简化,同时也使得功耗与传统结构相比有明显的降低。本文将对整体电路基本工作原理、比较器降低回馈噪声设计、采样设计和低功耗设计分别进行分析和介绍。

2 SAR ADC 结构及基本原理

2.1 系统结构与原理

本设计使用全差分输入结构,内置 DAC 由全电容阵列构成,对输入信号进行电容上极板采样。整体电路结构如图 1,包括比较器、采样开关、电容网络和连续逼近数字控制逻辑。这种结构可以有效的抑制衬底噪声和电源噪声,并能够提供良好的共模噪声抑制。DAC 使用二进制加权电容代替 $C-2C$ 电容来获得更好的线性度和匹配对称度。在这个基于电荷重新分配的结构中,电容网络能够提供采样保持和实现内置 DAC 功能。在本文提出的 ADC 结构中,输入信号通过栅压自举开关被采样到电容阵列的上极板上进行保持,通过栅压自举开关采样可以增加建立速度和输入带宽。电路工作开始时,首先采样开关导通进行采样,所有电容下极板接 V_{ref} 。下一个时刻,采样开关关闭,在没有任何电容开关动作时,比较器得到第一个比较结果。根据比较器的比较结果,使得接比较器输出为 1 的差分一侧的 MSB 电容接地,而另一侧的不变。ADC 重复这个步骤直到 LSB 位输出比较结果。在每个循环中,只有一个电容开关翻转,这样可以降低内置 DAC 网络对比较器的影响和减少开关翻转次数,从而降低功耗。传统结构和本文提出的结构之间,主要差别是:本文提出的结构开关顺序中没有向上的转换。在相同的晶体管尺寸下,NMOS 的导通电阻只有 PMOS 的 $1/3$,没有向上的转换,从而加快 DAC 的建立速度。此外,因为是采样到电容的上极板,比较器可以在没有任何电容开关动作的情况下直接进行第一次比较。因此对于 n 位 ADC,电容阵列中的单位电容是 2^{n-1} ,只有传统结构的一半。

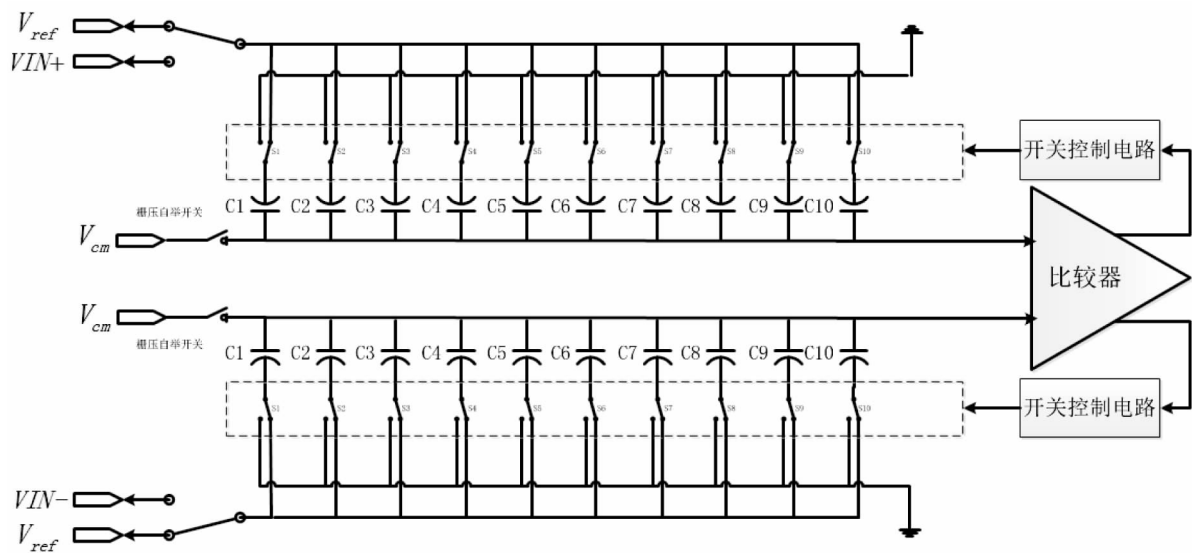


图 1 提出的电路整体结构

图 2 为一个运用本文所提结构的 3 位转换示意图。图中展示了转换过程中的所有可能性。在每一次比较过程中比较器会判断差分输入两侧的大小,然后内置 DAC 根据比较器输出结果控制对应一侧的电容开关接地,以使得其电压减去对应的值。图 2 中标示出了每个转换过程中开关相位所消耗的能量。传统的开关顺序在所有的尝试都成功的情况下消耗能量最小,然而当尝试不成功传统结构的开关顺序会有返回动作导致消耗多余的能量。提出的结构是在采样开关关闭后,比较器进行第一次比

较,在比较出结果之后开始有开关动作。因此,提出的方案中第一次比较中不需要消耗任何开关能量。与此相对比,传统的开关顺序在第一次比较之前会消耗能量。此外,在后面的开关顺序中,本文提出的理论也比传统的理论更加有效率。对于 n 位传统 SAR ADC,如果每个数字输出代码是等概率的,则平均开关能量为

$$E_{avg,conv} = \sum_{i=1}^n 2^{n+1-2i} (2^i - 1) CV_{ref}^2$$

使用提出的方案的开关平均能量为

$$E_{avg,mono} = \sum_{i=1}^{n-1} (2^{n-2-i}) CV_{ref}^2$$

对于 10 位来说,传统的开关消耗 $1365.3 CV_{ref}^2$,然而提出的方案只消耗 $255.5 CV_{ref}^2$ 。提出的技术在这项上就比传统的结构减少开关能耗 81%。

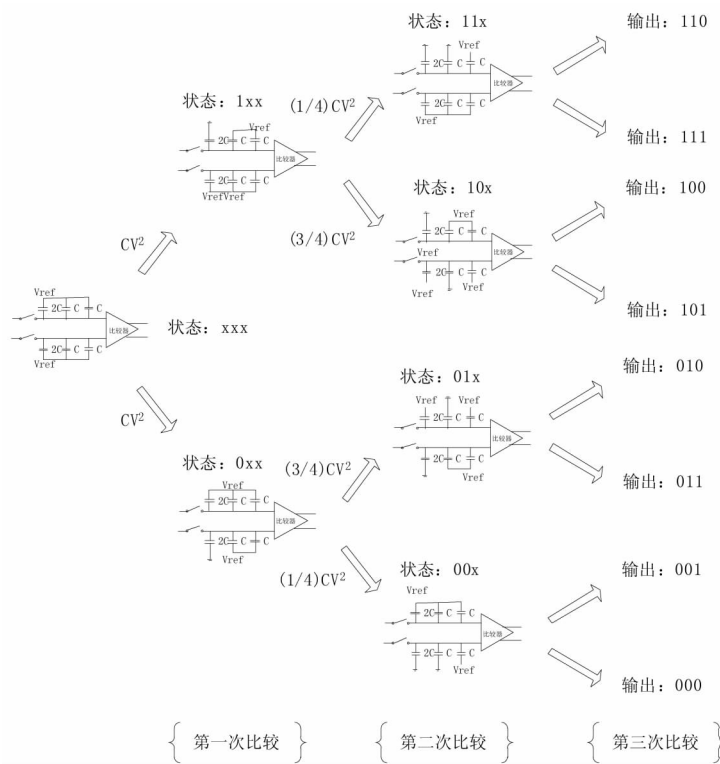


图 2 提出的线性开关逻辑

2.2 低回馈噪声比较器设计

由于比较器的输入端直接与 DAC 输出端相连,因此比较器的输入端会受到 DAC 输出端的影响,为了减小这种噪声的影响,本文使用了如图 3 的结构。相比传统结构增加了 M3, M4, M10, M11 四个晶体管。首先考虑没有这四个晶体管的情况下,当 clk 低电平时, V_X, V_Y 点电压接 VDD, 当 clk 高电平时接 GND, 这样 V_X 和 V_Y 会形成一个大的电压差, 这个电压差会通过输入管影响到比较器的栅极输入。对于本文使用的比较器结构, 当 clk 是低电平时 V_P, V_Q, V_X, V_Y 都接 VDD。当 clk 是高电平时 M3, M4 关断使得 V_P, V_Q 与 V_X, V_Y 分离, M10 和 M11 打开使得 V_P, V_Q 继续维持在 VDD 电位。因此在 M1 和 M2 源端没有电压变化, 从而减小了回馈噪声。

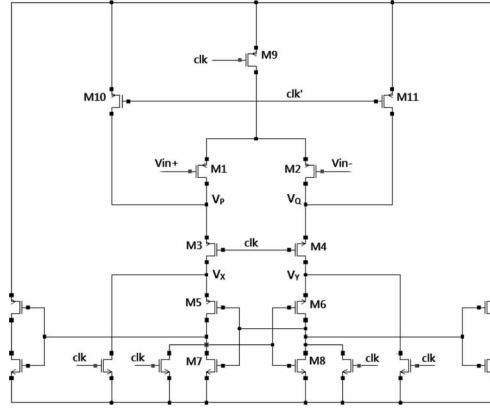


图 3 比较器结构

2.3 电容阵列排布

本结构中的电容阵列,不要求各个电容的容值的绝对精度,但对各个位权电容的相对精度有很高的要求,因为集成电容大部分都有 $\pm 20\% \sim \pm 30\%$ 的误差,所以模拟集成电路的精度以及性能一般都由器件匹配获得。系统的失配主要由于以下几点:工艺的偏差、接触电阻、电流不均匀流动、扩散相互影响、机械应力、温度梯度等。设计匹配器件是为了使它们对引起系统偏差的各种原因不敏感。两个器件之间的失配常表述为测得的器件与预期比率之间的偏离。如果预期值分别是 X_1 和 X_2 ,测量值分别为 x_1 和 x_2 ,则失配为

$$\delta = \frac{\left(\frac{x_1}{x_2}\right) - \left(\frac{X_1}{X_2}\right)}{\left(\frac{X_1}{X_2}\right)} = \frac{X_1 x_2}{X_2 x_1} - 1 \quad (1)$$

所有器件在尺寸及构成上会呈现出微观的不规则性,一种发生在器件边缘,称为边变化;另一类发生在整个器件上,称为面变化。由于不规则性与器件面积成比例,所以大部分器件的匹配取决于后者。

根据统计理论推知,面变化可以建模为:

$$s = m \sqrt{\frac{k}{2A}} \quad (2)$$

其中的 m 和 s 分别为有源面积为 A 的器件的某参数的平均值和标准差。 k 为匹配系数,该系数的大小取决于失配源。匹配系数和器件本身有关,不同类型器件的匹配系数相差很大。忽略系统误差,电容器的平均值等于电容 C ,并且正比于器件的面积。基于这两个因素,式(2)化简为

$$s = \sqrt{\frac{k_c C}{2}} \quad (3)$$

其中 k_c 是电容失配系数,具有电容的单位。计算任意两电容之间的失配均可使用该式。比如对于电容值均为 C 的两个匹配电容,把(3)代入(2)可得到两电容之间失配的标准偏差:

$$s_\delta = \sqrt{\frac{k_c}{C}} \quad (4)$$

上式说明了电容匹配和电容尺寸之间的基本关系,相当精确的匹配电容只能通过校正得到。对于电容值可能不相等的任意电容 C_1 和 C_2 间的失配为:

$$s_\delta = \sqrt{\frac{k_c (C_1 C_2)}{2 C_1 C_2}} \quad (5)$$

由式(5)可以看出,匹配电容中较小的电容对失配起着主要作用,也说明不宜采用大电容比例,所以只要可能,电路中应该避免使用大比例的匹配电容。可以通过采用器件串联构造小电容的方法避开这个问题。

工艺偏差对电容的影响,表现为引入系统失配,但是当匹配电容的面积一周长比相等时,则几乎没有影响。所以,通常会把相同的匹配电容设置成正方形。但是大电容需要绘制成矩形,矩形电容会增加周长波动而引起失配,此时可以采用匹配值电容或者单位电容阵列。

精确匹配电容会很容易受到导线寄生电容的影响,导致系统失配,如果影响很大,可以调整各导线长度,使它们的比率和相应电容的比率匹配。

因此构造匹配电容器的基本原则主要有:匹配电容尽量是相同图形的电容,并且精确匹配电容采用正方形;匹配电容要相邻摆放,在场氧化层上摆放匹配电容;将高阻节点连接到匹配电容的上极板;在阵列的外围放置陪衬电容;对匹配电容进行静电屏蔽,不要将线走在没有进行静电屏蔽的电容上;尽量在低应力梯度区域放置电容;匹配电容应远离功率器件;精确匹配电容应沿着管芯对称轴放置。

3 电路后仿真

基于 SMIC 0.18um 1P6M CMOS 工艺对此 SAR ADC 进行设计,电容阵列采用金属电容以确保更好的匹配性,本设计采用 spectre 进行仿真分析,设计参数为分辨率 10 位,采样 10Ms/s。全差分输入信号为 1Ms/s,输入范围 0~1V,电源电压 1.8V。芯片的整体版图如图 4。对芯片进行后仿真,对结果进行 FFT 频谱分析得到图 5 所示结果,可以看到 SNDR 达到 59dB,即有效位数达到 9.5 位。图 6 为不同输入信号频率下的有效位数变化情况,可以看到在接近奈奎斯特采样频率情况下有效位数仍然达到 9 位。

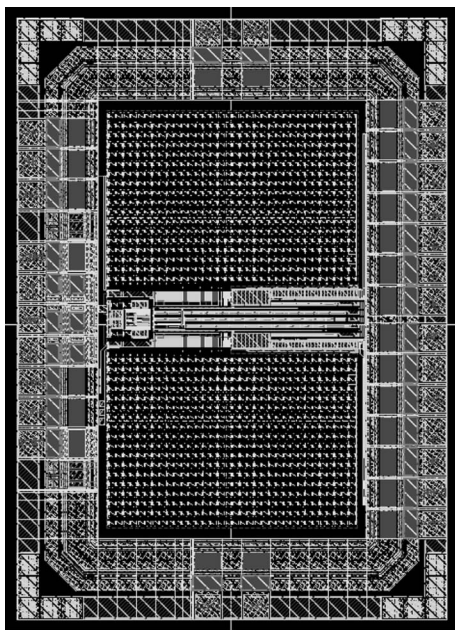


图 4 SAR ADC 整体版图

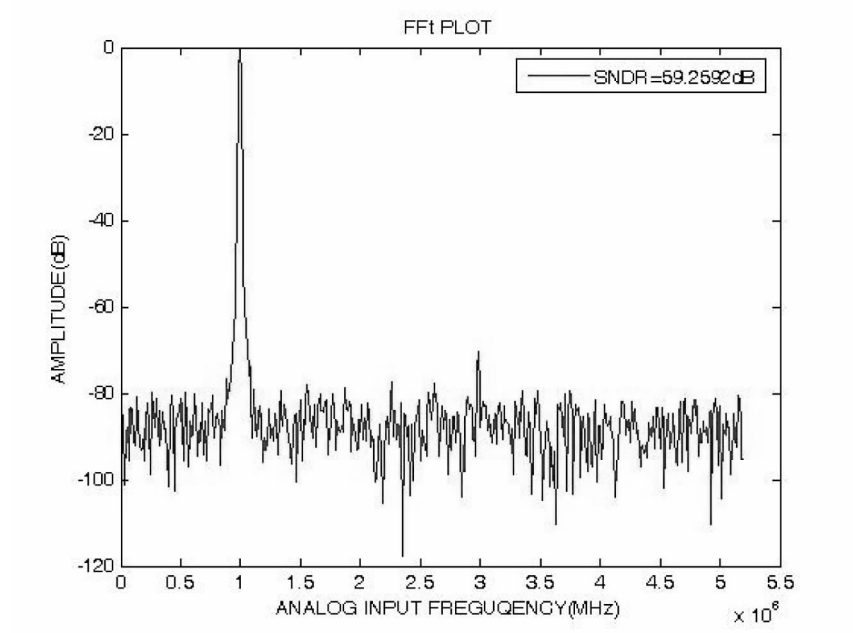


图 5 输出信号频谱图

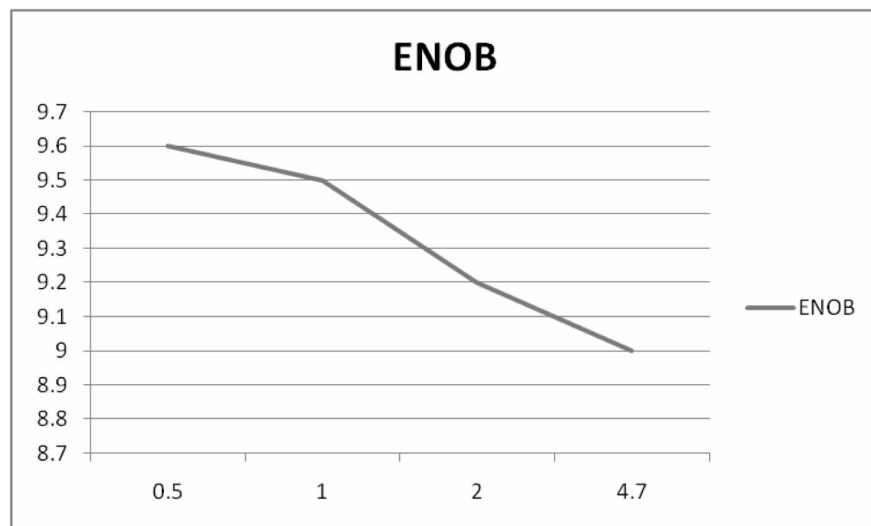


图 6 不同输入信号频率下有效位数

4 总结

本文设计的全差分 10bit 低功耗逐次逼近转换器芯片,在 0.18 μm CMOS 工艺下,所占用的面积为 0.6 mm^2 ,电源电压为 1.8V。本文采用的上极板采样结构新颖,能使得逻辑简化和降低功耗(可以简化逻辑和降低功耗),速度也有进一步的提升。比较器使用降低回馈噪声设计,对电容阵列的排布进行了优化从而有效提高匹配性。后仿真结果有效位数达到 9.5 位。本模数转换器可用于对速度和精度不高,但却要求低功耗低电压的电子仪器设备中。

参 考 文 献

- [1] Liu C C, Chang S J, Huang G Y, et al. A 10-bit 50-MS/s SAR ADC With a Monotonic Capacitor Switching Procedure[J]. Solid-State Circuits, IEEE Journal of, 2010, 45(4):731 - 740
- [2] Zhang H, Qin Y, Yang S, et al. Design of an ultra-low power SAR ADC for biomedical applications[C]. Solid-State and Integrated Circuit Technology (ICSICT), 2010 10th IEEE International Conference on. IEEE, 2010: 460 - 462
- [3] 模拟电路版图的艺术:第二版/(美)黑斯廷斯(hastings, A.)著;张为等译[M];电子工业出版社,2007
- [4] Liu C C, Chang S J, Huang G Y, et al. A 0.92mW 10-bit 50-MS/s SAR ADC in 0.13 μ m CMOS process[C]. VLSI Circuits, 2009 Symposium on. IEEE, 2009: 236 - 237
- [5] Draxelmayr D. A 6b 600MHz 10mW ADC array in digital 90nm CMOS[C]. Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International. IEEE, 2004:264 - 527 Vol.1

Design of 10bit 10MHz SAR ADC

keqing ning, liyao shi, lan dai

(School of Electronic and Information Engineering, North China Univ. of Tech. , Beijing, 100144)

Abstract Due to the merits of low power and low area consumption successive approximation register analog-to-digital converter (SAR ADC) is popular application in electronics system. In this paper it presents a 10 bit 10-MS/s successive approximation register analog-to-digital converter that use a monotonic capacitor switching procedure, a new layout algorithm is adopted to reduce the mismatch in the capacitor array. It reduces the power by Signal sampling on the plus of capacitor and optimal capacitor switching logic. A low kick-back noise latch is proposed in the design of comparator. At last, it optimizes the parasitic parameters of each part circuits in the layout of SAR ADC. The chip was fabricated using SMIC 0.18 μ m CMOS technology with 1.8V supply, the 10 bits 10 MS/s SAR ADC achieves SNDR of 59.0dB and the ENOB IS 9.5 bits in the post-simulation. The ADC core occupies an active area of 0.6mm².

Key words SAR ADC; comparator; capacitor array

可编程增益放大器的分析与设计

研究生 朱江南

导师 杨 兵

(北方工业大学电子信息工程学院 北京 100144)

摘 要 可编程增益放大器是集成电路中的重要模块,广泛地应用在无线通讯系统、医疗等领域。可编程增益放大器是自动增益控制环路中的重要组成部分,其作用是后面电路提供稳定的输出信号幅度,防止电路因为输入信号太大而发生饱和或因为输入信号太小而不能正常工作,这样有利于后面电路发挥最优的系统性能。本文从自动增益控制环路基本原理开始,介绍了可编程增益放大器的基本工作原理和一些设计指标。经过对可编程增益放大器详细的理论分析,设计了一款闭环结构的可编程增益放大器。设计采用 TSMC 0.18 μm CMOS 工艺实现,仿真结果表明性能指标满足预期要求。

关 键 词 可编程增益放大器;自动增益控制;闭环

作者简介 朱江南,男,1989 年出生,硕士研究生,主要研究方向为模拟集成电路设计。

E-mail: 929143371@qq.com

1 绪论

随着科学技术的飞速发展,半导体产业不断地对集成电路提出更高且近乎苛刻的指标要求。虽然当下数字集成电路产品在数量上远远的大于模拟集成电路产品,而且数字集成电路已经能够处理越来越多的信号,但是模拟电路作为现实世界与数字电路之间的纽带,在许多复杂高性能的电路系统中模拟电路的存在被证明是必不可少的。

可编程增益放大器(Programmable Gain Amplifier)作用是对输入的模拟信号进行放大或者缩小。在数字电视调谐器^[1]、电磁能量测量器^[2]、磁数据存储系统^[3]、磁盘驱动器读通道^[4]等多处都有应用。

可编程增益放大器最显著的优点就是它可以使系统发挥最佳性能。我们知道,模拟信号要经过模数转换器(Analog to Digital Convert)最终变成数字信号,而模数转换器的最大信噪比是在一个特定的输入幅值下达到的。可编程增益放大器连接在模数转换器之前可以把模数转换器的输入信号也就是可编程增益放大器的输出信号的幅值稳定在这一特定值,从而达到模数转换器的最大信噪比。

早期的可编程增益放大器大多使用双极工艺来实现,其一是因为当时模拟系统都采用的是此工艺,其二是因为双极工艺器件本身具有的指数特性很方便可编程增益放大器进行设计。但是随着时间的推移,系统集成度的日益增高,人们希望把模拟和数字的功能都集中在一个芯片上实现,这样的话即节省了芯片面积又提升了系统的稳定性,这时候 CMOS 工艺具有的低成本特点逐渐进入了人们的视野。但是 CMOS 工艺器件的平方律特性使得可编程增益放大器的设计难度加大,而且 CMOS 器件的工艺复杂、工作速度较慢,所以我们需要采用新的思路来进行可编程增益放大器的设计。因此采用 CMOS 工艺实现的可编程增益放大器在当今时代有着积极且重要的意义。

本论文专注于可编程增益放大器(Programmable Gain Amplifier)的分析与设计。在总结原有结构的基础上,完成了从设计指标的确定、电路结构的选择及设计指标实现等一系列的过程。

2 可编程增益放大器概述

2.1 可编程增益放大器基本工作原理

由于外界输入的模拟信号大小未知,为了保证输出模拟信号的大小保持不变,可编程增益放大器的增益需要根据输入信号的变化来调整大小。自动增益控制(Automatic Gain Control)电路实现了增益随输入信号变化而调整的功能。

自动增益控制环路的核心部分是可编程增益放大器。如图 1 所示,可编程增益放大器的输出连接一个峰值检测电路,峰值检测电路的输出与一个参考电压 V_{ref} 相比较,两者比较完的结果通过滤波器后作为可编程增益放大器的控制信号,这三个模块形成了一个环路。从反馈分析可知,只要自动增益控制环路的增益很大,那么 V_{out} 就近似等于 V_{ref} 的值。

绝大多数的可编程增益放大器的增益和它的控制信号之间不是线性关系而是 dB 线性关系。可以从两个方面来解释:第一,在众多系统中输入的信号幅度范围较大,从而需要使得可编程增益放大器的增益变化范围很宽,如果增益与增益控制信号是线性关系,那么增益控制信号的变化范围就很大,太大的变化范围难以保证其控制精度。第二,增益控制信号和增益 dB 线性相关保证了增益控制环路的稳定时间与可编程增益放大器的增益无关。

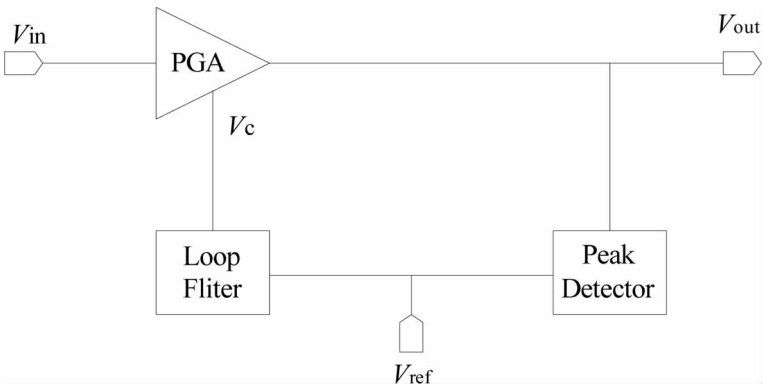


图 1 自动增益控制环路

在实际的设计中,自动增益控制环路的稳定时间太短或太长都不是我们所希望的。如果稳定时间太短,载波频率会发生变化,会使得某些载波系统发生解调错误。如果稳定时间过长,当输入信号变大时自动增益控制环路还没把增益降低下来,就会使得输出信号发生失真。

2.2 可编程增益放大器的参数指标

不同的系统在不同场合所应用的可编程增益放大器所需要的性能指标也不尽相同,下面是几个主要的可编程增益放大器的参数指标。

增益控制范围:这是可编程增益放大器和普通运放区别的一个重要参数,可编程增益放大器的增益是由增益控制信号控制的,增益控制范围越大系统的可输入动态范围也就越大。增益范围从几 dB 到几十 dB 不等,这主要取决于 PGA 的应用场所。

增益步长:增益控制信号变到相邻控制位时使增益发生改变的值。增益步长越小越有利于发挥模数转换器的最优性能,通常为 1dB。

增益带宽:也可用工作频率描述。由于可编程增益放大器工作时增益经常根据输入信号幅度大小而发生变化,所以运放的带宽也经常发生改变,为了使系统正常运转,最低带宽也能满足系统正常工作。

增益线性度:自动增益控制电路要求增益与控制信号之间成 dB 线性关系。一般要求它小于最小步长的二分之一。

失真:对当今的信号处理系统来讲,谐波失真和模数转换器精度存在有一定的联系。一般的系统会对信号的失真有一定的限制。

动态范围:通常指信号幅值的变化范围。通常用 1dB 压缩点和三阶交调点来描述信号的线性度。

输出直流失调:通常也描述为输出的共模电平大小不等,主要由于器件在生产制造过程中的不匹配造成。

这几个指标正如模拟电路设计的八边形法则那样既相互联系也相互制约。我们的设计目标通常是高增益控制范围、低增益步长、高带宽、高线性度和低功耗等。怎么样处理好各指标之间的关系就成为了设计过程中要处理好的主要问题。处理好各个参数间的关系将大幅地提高电路性能并推动模拟 IC 设计的发展。

3 可编程增益放大器结构

四种基本的增益控制方式如图 2 所示:

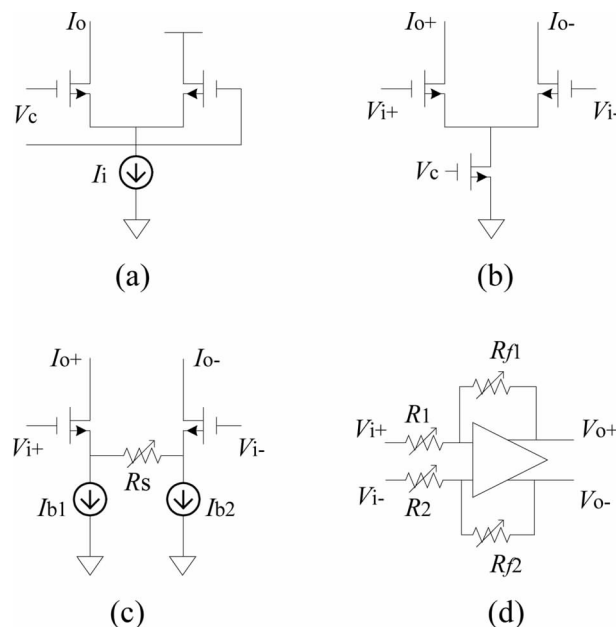


图 2 增益变化基本结构

图(a)是一个分流器。分流比率是由控制电压 V_c 决定。分流器的二阶特性使实现 dB 线性的增益控制很困难。需要对 V_c 进行一级预校正^[5]。整体的线性度被产生电流 I_i 的输入跨导制约。

图(b)中的源级耦合对跨导的改变是通过改变晶体管的偏置电流来实现的^[6]。电路的增益和输入

相关噪声分别和输入管的 g_m 和 $\sqrt{\frac{1}{g_m}}$ 成比例。当输入信号很小时,大的偏置电流用于实现高增益和低噪声。另一方面,当输入信号很大时,低偏置电流会降低电路的线性度。

在图(c)中,可以通过改变源级退化电阻 R_s 来改变源级耦合对的跨导。当输入信号很小时,较小的电阻 R_s 用来实现高增益和低噪声。当输入信号较大时,大的电阻 R_s 用来实现低增益和高线性度。因此,这种结构输出电平稳定,在不考虑增益多少的情况下,直接实现恒定的信号对噪声和失真的比率。

图(d)为一带电阻反馈的高增益放大器。它增益的改变可通过调节电阻 R_{f1}/R_1 和 R_{f2}/R_2 的比率来实现。当环路增益很大并且电阻网络呈线性时电路的高线性度可以实现。然而,如果采用常规的放大器^[7],那么反馈系数的改变会导致带宽和总谐波失真的变化。当电路在整个增益范围涵盖最坏的情况时,它的功耗不会是最优的。

实现可编程增益放大器电路的方式一般分为两种:一种为开环结构,一种为闭环结构。带负反馈的闭环结构对增益的准确度有良好的控制并且具有比较好的线性度。但是,在带宽方面,如果为了实现比较宽的工作频率范围则需要比较大的功耗来实现。相反,开环结构则更适用于低功耗和宽带的应用,但是开环电路的线性度很大程度上依赖于输入级的固有线性度,所以经常用到线性优化的办法来提高线性度。

4 可编程增益放大器设计与实现

4.1 指标分析及结构选择

经过对可编程增益放大器的工作原理以及结构分析,设计了一款具有 40dB 增益调节范围 1dB 增益步长的可编程增益放大器,它的性能指标如下:电源电压为 5V,工作频率为 100Hz,输入信号范围为 1mV 至 100mV,输出小信号幅值为 1.5V,可计算得出增益范围为 23.52dB 至 63.52dB,共模电压为 2V,增益误差小于模数转换器最小精度的一半。

根据上述要求可以看出,输入信号为低频信号,且动态范围较大,因此采用闭环结构来设计电路。

闭环结构电路由电阻反馈和电压放大器构成。由于增益变化范围较大,所以采用了三级级联的方式。第一级为固定增益级,实现 23.52dB 的固定增益,第二级为粗调级,实现从 0dB 至 36dB 以 6dB 为步幅的增益调节方式。第三级为细调级,实现从 0dB 至 5dB 以 1dB 为步幅的增益调节方式。这样整体就能实现从 23.52dB 至 63.52dB 以 1dB 为步幅的增益调节方式。整个放大器的结构框图如图 3 所示。

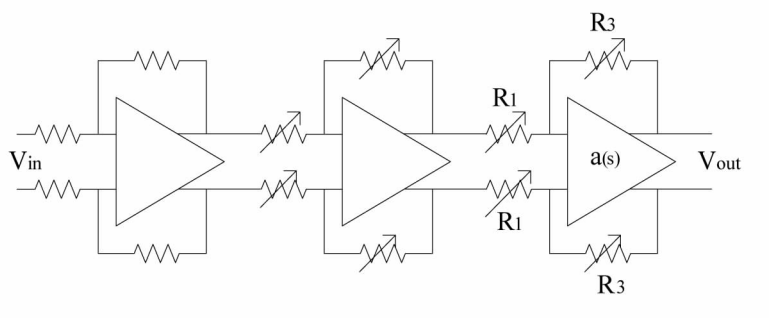


图 3 可编程增益放大器结构框图

$$20\log A_{closed} = \frac{40x}{N\ln 10} + \frac{40x^3}{3N^3\ln 10} \dots \quad (2)$$

在式(2)中,只要 N 相对较大, x 相对较小,那么等式右侧的第二项的值可以忽略,此时运放的闭环增益和 x 的值呈线性关系。在此次设计中我们令 $N=18, x=5$ 。提取式(2)中的线性项:

$$\frac{40x}{N\ln 10} \approx 17.37 \frac{x}{N} \quad (3)$$

在设计电路过程中由于 x 的值一定不大于 N 的值,所以这个电路最多能实现 17.37dB 增益。上面粗调级所要实现的最大增益为 36dB,粗调级的电路设计不能采用这个电路。

4.2 运算放大器的参数计算

全差分运算放大器在当今电路设计中被广泛的应用,相对于单端放大器,全差分放大器具有提供较大的输出信号摆幅,而且受共模噪声影响小等等优点。在一个完全对称的差分电路中,它的输出不存在偶次谐波的非线性。但全差分放大器也有一些缺点,其中之一就是它需要完全对称的反馈网络和一个共模反馈电路来控制输出共模电压使其输出共模电压稳定。因此谐波失真要求大于 60dB。我们以最后一级为例阐述放大器指标的制定。设运放的一阶响应为 $a(s)$, A_0 为低频增益, p_0 为其第一个主极点。

$$a(s) = \frac{A_0}{1 + \frac{s}{p_0}} \quad (4)$$

由于运放的输入虚地点电流相等,可知最后一级闭环增益为:

$$A(s) = \frac{-\frac{R_3}{R_1}}{1 + \frac{1}{A_0} + \frac{s}{A_0 p_0 \beta}} = \frac{-\frac{R_3}{R_1}}{1 + \frac{1}{A_0 \beta} + \frac{s}{\omega_u \beta}} \quad (5)$$

其中 $A_0 p_0 = \omega_u$, ω_u 为运放的单位增益带宽, $\beta = R_1 / (R_1 + R_3)$, β 为反馈因子。从上式可以得出闭环运放的增益为 R_3 / R_1 , 带宽为 $\beta \omega_u$ 。

根据式(5)可以计算出运放的指标。当增益为 36dB 时,也就是 $R_3 / R_1 = 64$, $\beta = R_1 / (R_1 + R_3) = 1/65$,为满足增益误差小于模数转换器最小精度的一半,闭环增益带宽大于信号带宽的 10 倍。由公式可计算出 A_0 和 ω_u 。

$$\frac{1}{A_0 \beta} \leq \frac{1}{2} LSB = \frac{1}{2} \frac{2.4}{2^{13}} \quad (6)$$

$$\beta \omega_u \geq 10 f_s = 10 \times 100 \quad (7)$$

由式(6)可以得出, $A_0 \geq 113\text{dB}$, $\omega_u = A_0 p_0 = 443733 \times 100 = 4437300$ 。由(7)式得出 $\omega_u = 6500$ 。因此为满足谐波失真需要满足 $\omega_u = 4437300$ 。

运放 A_0 需要满足大于 113dB 的增益,所以运放采用带增益提高结构的折叠式共源共栅电路,下面为具体的电路设计。

4.3 运算放大器的电路设计

此次设计的运放由三个全差分折叠式共源共栅放大器构成。其中一个作为主运放,另外两个作为辅助运放。完整的电路结构如图 6 所示。

由于 PMOS 管的噪声性能比 NMOS 管好,我们选择 PMOS 差分对作为主运放的输入级。辅助运放输入级的选择是根据它们的共模输入电压要求来确定的。辅助运放 A_1 工作在高共模电压下,其输入级为 NMOS 差分对;同理可知,辅助运放 A_2 输入级为 PMOS 差分对。辅助运放 A_1 的具体电路结

构如图 7 所示,辅助运放 A_2 的具体电路结构均与 A_1 相似只是输入级由 NMOS 变成 PMOS 差分对。

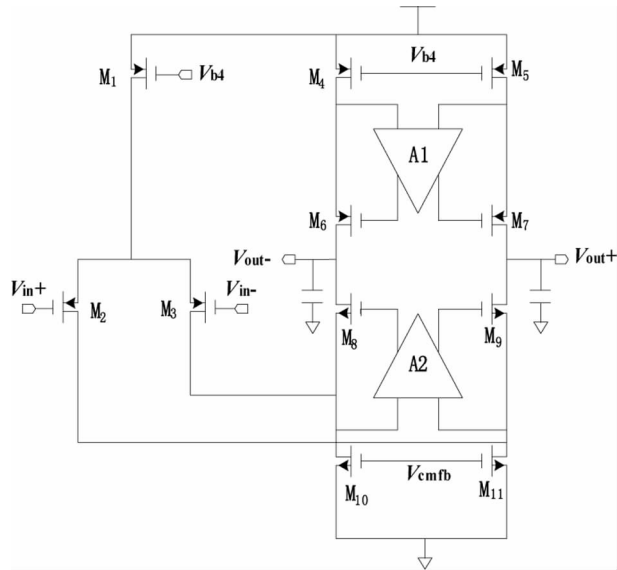


图 6 主运算放大器结构

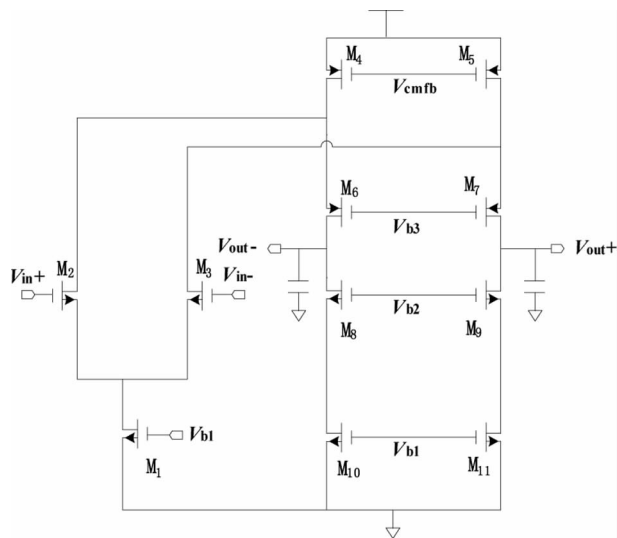


图 7 辅助运算放大器 A1 电路

随着最小沟道长度的越来越短,晶体管的本征输出阻抗也越来越小,从而运放达到预期的增益变得越来越困难。这里采用共源共栅电流镜提供偏置。然而,传统结构的共源共栅电流镜限制了信号摆幅,从而限制了在一些特定环境下的应用。所以本次设计采用宽摆幅电流镜为电路提供偏置。电路如图 8 所示:

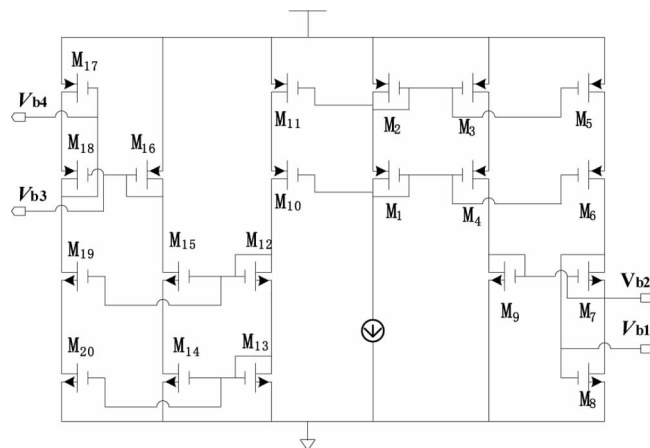
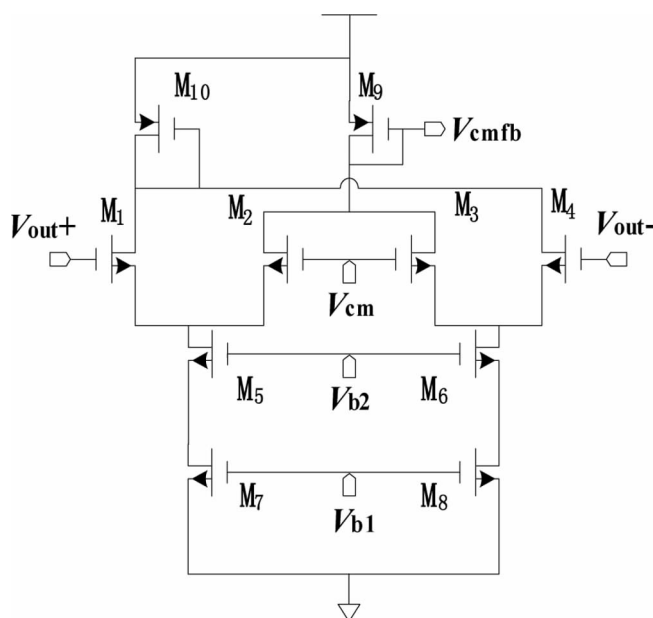


图 8 宽摆幅电流镜偏置电路

虽然全差分运算放大器电路具有更大的输出摆幅、对环境噪声有较强的抗干扰能力、避免了镜像极点等有很多优点。但是,高增益全差分电路必须要有共模反馈电路。共模反馈电路一般分为两类:连续时间的共模反馈和开关电容共模反馈。本次所设计的运放均采用连续时间的共模反馈,电路如图 9 所示,连续时间的共模反馈电路和开关电容共模反馈电路相比较,前者更快达到稳定输出。



如图 9 共模反馈电路

5 可编程增益放大器仿真结果

5.1 运放的仿真结果与分析

下图为运放的增益和频率响应的仿真结果图,从图中可以看出此运放增益 113.8dB,相位裕度为 67° 。-3dB 带宽为 282Hz,满足开始设定的运放设计指标。

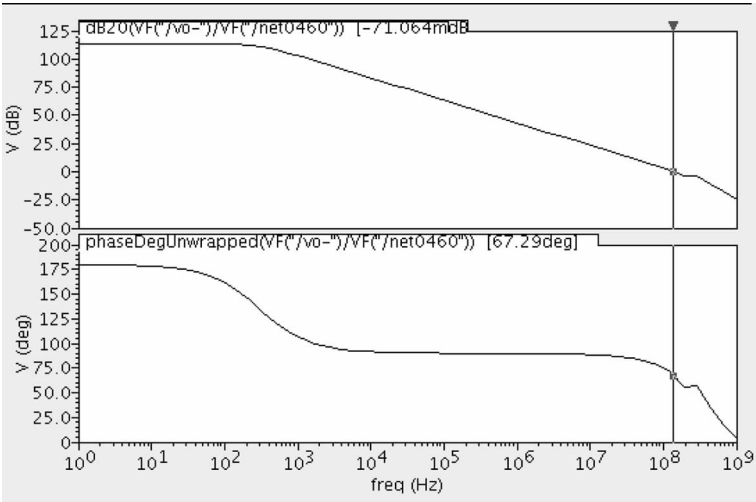


图 10 运放的增益和相位裕度

流经此放大器的静态电流为 $337\mu\text{A}$, 功耗为 1.685mW 。由于此结构的运放驱动电阻负载能力较弱, 所以在运放的输出端加入了两个源级跟随器用来增强运放驱动电阻负载的能力。仿真结果表明, 加入的源级跟随器具有较大的静态电流, 在未来的工作中如果能够采用具有较低功耗的驱动电路则可以进一步的降低电路功耗。

5.2 可编程增益放大器的仿真结果分析

表 1 PGA 仿真结果

理想值 (dB)	仿真值 (dB)	误差值 (dB)	理想值 (dB)	仿真值 (dB)	误差值 (dB)	理想值 (dB)	仿真值 (dB)	误差值 (dB)
23.522	23.577	0.055	37.522	37.515	-0.007	51.522	51.533	0.011
24.522	24.543	0.021	38.522	38.499	-0.023	52.522	52.563	0.041
25.522	25.515	-0.007	39.522	39.503	-0.019	53.522	53.626	0.104
26.522	26.499	-0.023	40.522	40.532	0.01	54.522	54.592	0.007
27.522	27.503	-0.019	41.522	41.59	0.068	55.522	55.564	0.042
28.522	28.533	0.011	42.522	42.556	0.034	56.522	56.549	0.027
29.522	29.57	0.048	43.522	43.528	0.006	57.522	57.552	0.03
30.522	30.536	0.014	44.522	44.513	-0.009	58.522	58.582	0.06
31.522	31.508	-0.014	45.522	45.516	-0.006	59.522	59.509	-0.013
32.522	32.492	-0.03	46.522	46.546	0.024	60.522	60.475	-0.047
33.522	33.496	-0.026	47.522	47.607	0.085	61.522	61.447	-0.075
34.522	34.526	0.004	48.522	48.573	0.051	62.522	62.431	-0.091
35.522	35.577	0.055	49.522	49.546	0.024	63.522	63.435	-0.087
36.522	36.543	0.021	50.522	50.53	0.008			

表 1 为此可编程增益放大器的仿真结果数据, 从此表格可以看出此可编程增益放大器的仿真结果

和预期结果误差很小,最大误差值为 0.104dB,主要误差原因来源于我们采用的反馈电阻均为整数倍的单位阻值,虽然整数倍的单位电阻在版图实现过程中能够达到很好的匹配,但是整数倍的电阻比值转换成 dB 表达时会存在一定的误差,但总体设计满足预定指标要求。

6 总结与展望

在仔细分析各种可变增益放大器结构以及一些增益控制技术的基础上,本文设计了一款具有 40dB 增益步长、1dB 增益步幅的可编程增益放大器,并采用 TSMC 0.18 μm CMOS 工艺实现。仿真结构满足预期指标。

运放的结构大体上可以分为两种:一种是开环结构,一种是闭环结构。开环结构的可变增益放大器一般是通过改变它的等效负载或改变它的等效输入跨导从而实现增益的变化。闭环结构一般采用改变它的反馈阻值从而实现运放增益的改变。闭环结构的运放和开环结构的运放相比较,闭环结构的电路工作会更加稳定并且对温度、工艺、电压的改变不是很敏感。但选择何种结构的运放还需要根据电路的工作环境来确定。开环结构适用于中、高频的应用。闭环结构适用于低频结构。

本文所设计的可编程增益放大器是基于改变反馈电阻阻值实现增益控制。所设计的电路的粗调级和细调级分别采用不同的电路连接方式以尽量降低增益误差。可编程增益放大器的运放根据具体的指标要求选择了带增益提高的折叠式共源共栅结构。

最后的仿真结果表明在 40dB 的增益控制范围内,最大增益误差为 0.102dB。和同类电路相比有较高的电路精度。主要的增益误差是由于所采用整数倍电阻阻值所引起的。可以采取微调电阻阻值的方法进一步减小增益误差,但电路都采用整数倍电阻会有利于版图的匹配和工艺制造过程。

展望未来工作可以继续对可编程增益放大器中的运放进行研究,对电路的其他性能进行进一步的优化,如降低温漂,降低低噪声和减小电路的功耗等等。

参 考 文 献

- [1] Youn Y S, Kim C S, Kim N S, et al. A 1 GHz band low-distortion up-converter with a liner in dB control VGA for digital TV tuner [A]. IEEE RFIC Dig Paper [C]. AZ, USA. 2001. 257—260
- [2] Wintenberg A L, Simpson M L, Young G R, et al. A CMOS variable amplifier for PHENIX electromagnetic calorimeter and RICH energy measurements [J]. IEEE Trans Nucl Sci, 1997, 44(3): 102—106
- [3] Gomez R, Abidi A A. A 50 MHz CMOS variable gain amplifier for magnetic data storage systems [J]. IEEE J Sol Sta Circ, 1992, 27(6): 935—939
- [4] Harjani R. A low-power CMOS VGA for 50 Mb/s disk drive read channels [J]. IEEE Trans Circ and Syst II, 1995, 42(6): 370—376
- [5] O. Watanabe, S. Otaka, M. Ashida, and T. Itakura, “A 380—MHz CMOS linear-in-dB signal-summing variable gain amplifier with gain compensation techniques for CDMA systems,” in Symp. VLSI Circuits Dig. Tech. Papers, 2002, pp. 136 - 139
- [6] P. Orsatti, F. Piazza, and Q. Huang, “A 71—MHz CMOS IF-baseband strip for GSM,”IEEEJ. Solid-State Circuits, vol. 31, pp. 104 - 108, Jan. 2000
- [7] J. Guido, V. Leung, J. Kenney, J. Trackim, A. Agrillo, E. Zimany, and R. Shariatdoust, “Analog front end

- IC for category I and II ADSL,” in Symp. VLSI Circuits Dig. Tech. Papers, 2000, pp. 178 – 181
- [8] Xiaobin Shen, Taotao Yan, Yuxiao Lu, Jiangjun Zhou, “A 0.25dB Gain Step High Linear Programmable Gain Amplifier”, IEEE ASIC (ASICON) IEEE International Conference, pp. 1062–1065, 2011

Design and Analysis of Programmable Gain Amplifier

Zhu Jiangnan

(School of Electronic and Information Engineering, North China Univ. of Tech. , Beijing, 100144)

Abstract Programmable Gain Amplifier(PGA) is an important module in the integrated circuits. It is widely used in wireless communication system, medical treatment and etc. PGA is an important part of Automatic Gain Control(AGC) loop, AGC loop offers constant signal amplitude for subsequent circuit to prevent that integrated circuits work saturated by large input signal or be irregularly by too weak input signal and helps the subsequent circuit to play better system performance. This paper research from the basic concept of AGC loop, and introduce the principle and some design index of PGA. Based on the detailed theoretical analysis of PGA, a PGA has been designed. Design has been achieved in TSMC 0.18 μm CMOS process, the simulation result shows the performance of the PGA meets the expected requirements.

Key words programmable gain amplifier; automatic gain control; closed loop

对智能冰箱及未来交互性设计发展的研究

研究生 祁文静

导师 王书万

(北方工业大学建筑与艺术学院 北京 100144)

摘 要 冰箱的发展随着技术的不断更新越来越趋向智能化,但是冰箱智能化的发展趋向还是局限于冰箱最为原始的功能上,而没有体现未来信息化时代的特征。本文认为,在信息化时代,冰箱与人之间的交互关系是冰箱真正走向智能化的临界点。尽管中国的电子商务实现了大多数商品的线上购物,但人们日常生活中食用的生鲜食品则由于不宜碰撞,保鲜期短等特点无法实现长距离的运输;但是,由于当今社会生活压力大,节奏快,很多年轻人又缺少足够的时间进行日常的生活用品补给,因此给人们带来了很大的不便。实现生鲜食品的线上购物可以使人们的生活更加便利。

本课题首先介绍了近几年国内传统零售市场以及电子商务市场的数据发展现状及分析,以及传统零售商所做的转型努力和效果分析。之后分析了冰箱智能化发展的现有成果,以及搭载无线网络以及电子触屏的可行性。基于以上的分析了解,最后针对智能冰箱的应用程序的界面设计进行了人群定位调查、功能预设和色彩应用原因的梳理等,由此确定了个人中心、搜索、菜谱、主界面、收藏、购物车、交流区等七个部分的界面设计和交互方式。界面交互的设计以及整体流程的梳理是本论文的核心内容。

关键词 智能冰箱;销售模式;零售;电子商务;交互设计

作者简介 祁文静,女,1989 年出生,硕士研究生,主要研究方向为工业设计。

E-mail: binfexumo@163.com

1 绪论

1.1 课题研究背景

1.1.1 电商近年发展分析

电商作为一种新兴的商业发展模式,在最近的十年间发展迅速,对传统的商业模式形成了巨大的冲击。

1.1.1.1 电商发展迅速的原因分析

电子商务的经济模式其实早已出现,在美国,亚马逊和 eBay 已发展多年,但是并未对其传统市场造成太大的冲击,究其原因是因为美国的传统市场模式发展完善,导致传统市场模式下的商品与电子商务中的商品在价格、质量等各方面相差无几;而在中国,由于各方面的限制和压力,导致传统市场模式下的商品与电子商务模式下的商品在价格方面有着明显的劣势。

1. 土地租金的问题

自全球金融危机以来,中国的经济发展受到影响,由此房地产行业受到非经济方面的因素影响,开始迅猛发展,导致中国的各个城市的商业用地租金上升,由此产生的商业成本大幅度提升,因此产品的价格也开始大幅度提升。而相对于传统市场,电子商务的市场则无需面临这方面的难题,基于虚拟网络为载体而实现的商品交易,本身就规避了实体商业模式中的土地租金这一环节。

2. 传统商业链的层层利润

传统的商业模式大多是由总经销商获得产品的销售权,再一层一层向下发展各个地区的经销商,经过中国六个大区后(东北、华北、中南、华东、西南、西北),再经省、市、区县等,每一层的成本和利润都会使得产品的价格向上提高一个阶梯段,由此最终产品的定价远超于产品本身的价值,而与此相对的电子商务则省去了中间经销商的成本,由此产品的价格更为贴近产品本身的价值,因此对于消费者来说无疑是更为实惠的选择。

3. 基本资金的需求和短缺

在中国的传统市场,商家想要创业的前期资金是重中之重,而这恰恰是初期创业的人的短板。资金的不足使得对产品的传播和经营都受到相当程度的限制,而初期创业的各方面资金投入又必不可少,像如房租、人员成本、基本硬件设施的配备等都要纳入考虑,在初期摸索发展方向和无法盈利的阶段,这些投入显然需要一定的资金储备。而电子商务则无需如此,在网络上开店的便捷以及近乎零成本的开店使得商家仅需将产品的成本专注在产品本身,而无需考虑其他的营运成本。中国有大量的企业,规模在微型企业以下的标准,在传统市场较难存活,也无法与大企业竞争,因此电子商务的平台给了这些企业机会,使得他们的产品可以面向消费者,获得可观的利润,因此没有资金储备的创业人员更多地会选择用电子商务的模式来推销商品,并在过程中积累经验,由此获得商业发展。

4. 对于商品的选择更为便捷直观

相对于传统模式中人们的购物只局限在一个或几个购物广场内,电子商务使得人们在购买某种商品时可以在全国范围内进行比价,挑选,并且不需外出,增加了商品的可选性,也更为节省时间,更为方便。

因此基于中国经济社会等各方面的现状,中国传统市场的不完善以及各种限制,使得大量的创业者和消费者渐渐更加重视电子商务的发展以及其产品,由此电子商务在近年来发展迅猛,抢占了传统市场的份额。

1.1.1.2 电子商务发展的现状

自1999年阿里巴巴成立以来,经过几年的时间,中国电子商务开始迅猛的蓬勃发展,从B2B模式到B2C和C2C模式,电子商务的平台越来越多,像如京东商城,1号店,天猫,淘宝,聚美优品等都已经慢慢成为人们的消费习惯。艾瑞统计数据显示,2014年中国电子商务市场交易规模12.3万亿元,增

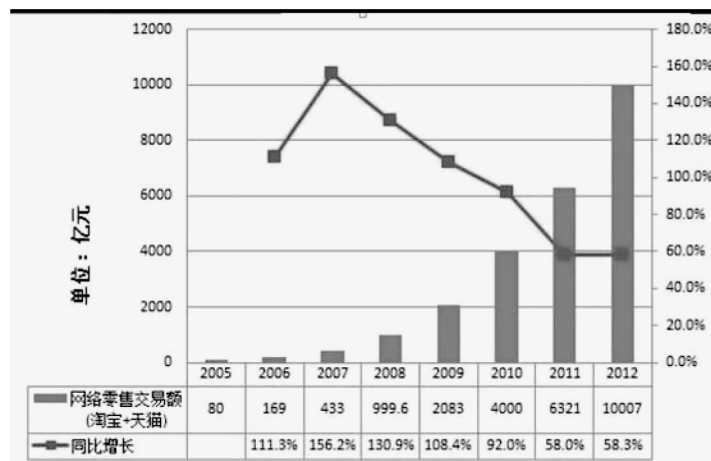


图 1-1

长 21.3%，其中网络购物增长 48.7%，在社会消费品零售总额渗透率年度首次突破 10%，成为推动电子商务市场发展的重要力量。另外，在线旅游增长 27.1%，本地生活服务 O2O 增长 42.8%，共同促进电子商务市场整体的快速增长。天猫和淘宝是最具代表性的电子商务平台，从这两家的发展规模以及近几年的营业额就可以明显地看出电子商务的发展速度。

图 1—1 是 2005 年到 2012 年的淘宝和天猫网络零售交易额，从 2005 年到 2010 年五年间其增长速度基本保持在百分之百，而在 2011 年到 2012 年由于京东商城、1 号店等 B2C 平台的出现，其增长速度有所下缓，但仍然保持在较高的水平，到 2012 年淘宝和天猫的销售总额已经突破 1 万亿。

	2012			2013				2014
	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
淘宝 GMV	167	179	255	223	257	275	346	295
天猫 GMV	42	49	91	71	88	99	183	135
总 GMV	209	228	346	294	345	374	529	430
移动端 GMV	4.6%	5.6%	7.4%	10.7%	12.0%	14.7%	19.7%	27.4%

来源：阿里巴巴 F1-A；单位：十亿人民币

制图：腾讯科技

图表 1—2(单位:十亿)

图 1—2 是自 2012 年第二季度到 2014 年第一季度之间各季度的交易总额，总体呈上升趋势，从移动端的迅速增长来看，消费者已经越来越习惯使用手机等移动终端来进行线上消费，改变人们的消费习惯，这本身对于传统的消费模式就是一种潜伏性而具有强大竞争力的冲击。

而淘宝和天猫在电子商务领域所占的份额也随着其它平台的出现而渐趋减少，图 1—3 是 2012 年第三季度 B2C 领域内各个商务平台的占比，可以看出天猫占整个 B2C 领域的 55% 左右，而其他的 45% 的市场份额则由京东商城、亚马逊中国、腾讯 B2C 等占据大部分。2012 年第三季度的天猫销售总额为 490 亿元，由此可见整个电子商务平台中仅 B2C 就已经占据了相当大的市场份额。2012 年 1—12 月份，全国 50 家重点大型零售企业商品零售额同比增长 10.3%，增速较 2011 年全年低 10.5 个百

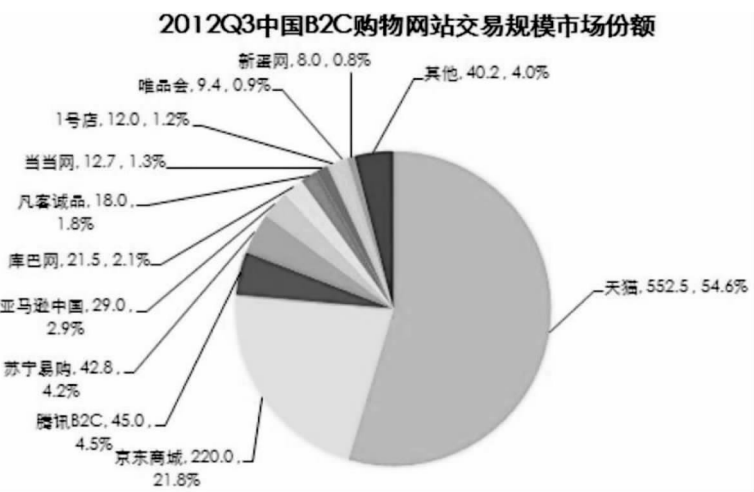


图 1—3 源自艾瑞数据

分点,销售增速放缓明显。(见表1-4)据联商网报道,其中,食品零售额同比增长14.2%,服装零售额同比增长10%,家电零售额同比增长4.5%,金银珠宝零售额同比增长10.9%,日用品零售额同比增长10.7%,化妆品零售额同比增长12.6%,各主要商品零售额同比增速较上年均有不同幅度的放缓。由此可见,传统零售模式的市场份额已经大规模地受到了电子商务模式的冲击。

业态	序号	企业名称	销售额(亿元)	销售额增幅	净利(亿元)	净利增幅	门店总数
百货	1	重庆百货	281.2	12.43%	6.99	15.59%	
	2	王府井百货	182.64	8.97%	6.7	15.60%	
	3	百盛集团	172.11	4.80%	8.5	-24.20%	54
	4	天虹商场	143.77	10.29%	5.88	2.40%	59
	5	银泰百货	139	19.80%	9.725	18.40%	
	6	万达百货★	111.8				
	7	茂业百货	110.67	6.10%	8.25	2%	39
	8	广百股份	73.49	2.30%	1.93	-9.52%	
	9	杭州大厦★	59.07	-2.80%			1
	10	中国春天百货	54.19	12.70%	1.56	-50.00%	16
	11	翠微股份	49.48	3.01%	1.48	12.62%	
	12	南京中央商场	40.7	15.40%			
	13	新世界百货	*30.604	20.22%	*4.891	-31.62%	41
	14	南京新百	28.5	55.09%	1.91	128.48%	
	15	杭州解百	21.46	-2.89%	0.774	0.78%	
	16	兰州民百	13.2	11.10%	0.597	26.96%	
	17	百大集团	12.25	-5.58%	0.87	10.45%	
	18	庄胜百货	2.09	-20.00%	1.65	18%	
超市	19	华润万家	941	13.78%			4425
	20	高鑫零售	778.51	14.30%	24.09	50.60%	
	21	家乐福中国	*459.88	10.80%			219
	22	联华超市	289.9	5.33%	3.4	-45.77%	4698
	23	永辉超市	246.84	39.21%	5.02	7.54%	
	24	麦德龙中国	*151.2	*23%			64
	25	人人乐	129.13	6.79%	-0.909	-153.65%	
	26	卜蜂莲花	106.77	6.60%	-3.92		74
	27	北京京客隆	91.84	6.40%	1.05	-50%	250
	28	超市发★	39.8	4.17%			
	29	红旗连锁	39.05	13.26%	1.73	1.69%	1336

表 1-4

1.1.1.2 电商发展模式的推进

从最开始流行于国外的 B2C 的商业模式,在中国发展出了阿里巴巴,京东商城,当当网等一系列公司,随后随着淘宝的出现,C2C 开始在中国盛行,成为本土发展出的一种电子商务模式。图 1-5 是艾瑞咨询出具的 2011—2013 年 B2C 和 C2C 在中国网络购物市场的规模占比,以及对未来 2014 年—2018 年的占比预测,数据显示,2014 年中国网络购物市场中 B2C 交易规模达 12882 亿元,在整体网络购物市场交易规模的比重达到 45.8%,较 2013 年的 40.4%增长了 5.4 个百分点。从增速来看,B2C 市场增长迅猛,2014 年中国网络购物 B2C 市场增长 68.7%,远高于 C2C 市场 35.2%的增速,B2C 市场将继续成为网络购物行业的主要推动力。而图 1-6 显示的是 2014 年中国 B2C 购物网站交易规模的市场份额分布情况,从 2014 年看,B2C 市场中,天猫市场份额占比超六成,京东占比为 18.6%,其余 B2C 企业中唯品会、1 号店、国美的增速均高于 B2C 市场整体增速。从自主销售为主 B2C 市场来看,京东占比近五成,苏宁易购占比达到 8.5%,唯品会占比达到 7.7%,其他像小米手机官网发展迅速,整

个市场集中度依然较高。

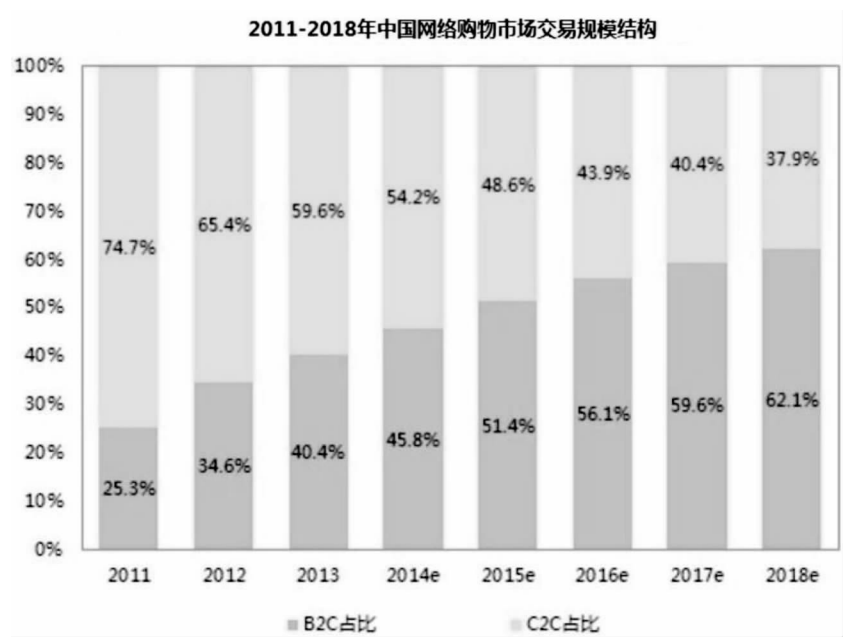


图 1—5 源自艾瑞数据

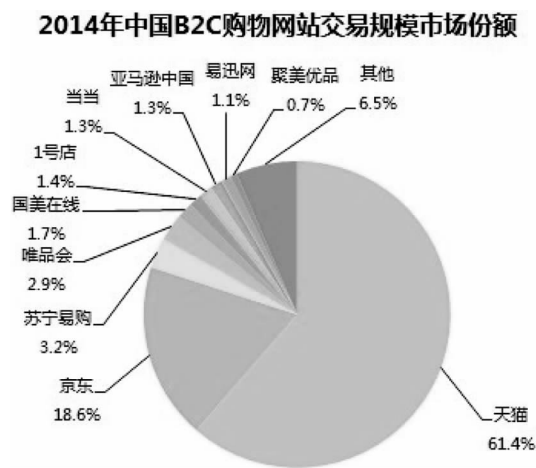


图 1—6 源自艾瑞数据

而传统百货公司为了应对电子商务的冲击,提出了电子商务所没有的购物体验的优势,由此出现了一种新的商业模式 O2O,即线下体验线上消费的一种模式,由此来凸显传统购物模式的优势。2014 年零售企业涉入了 O2O 的爆发期,大量的企业开始试水 O2O 模式,零售商诸如大润发、沃尔玛、苏宁、银泰等都开始进行线下和线上互为臂助的运营,以北京朝阳区大悦城为例,2014 年 11 月朝阳大悦城 O2O 平台开始上线,仅三周 APP 下载量已经超过了 1.5 万。在室内导航、寻车、会员等服务功能的基础上,以餐饮为核心,建立了购物中心首个集合外卖、团购、排号、订餐、点餐付款在内的餐饮全功能解决系统。

而与此对应的服务商也开始蓬勃发展,像如腾讯 POS 机硬件设备面世后,2014 年 6 月,腾讯与新世界百货、王府井百货、上品折扣等商业机构合作推广“微乐付”;同月时间里,微信推出免费 WiFi,此

次与微信合作的 wifi 提供商为 WiFiSong;同年六月,微信钱包中上线刷卡功能。自 9 月 15 日起,微信刷卡功能将支持 9 家连锁商店:DQ 冰雪皇后、国大药房、天虹、壹加壹、卜蜂莲花(华南)、好邻居、爱婴室、民生百货、国大 365。此外,像如支付宝,Paypal,苹果公司等针对 O2O 的营运模式也进行了相应的技术改变和运营调整。^[1]

由此可见,电商在近十年的发展规模已经对传统市场的运营规模造成了巨大的冲击,传统的商业模式像如百货等有着体验的优势而通过 O2O 的模式来予以应对,但大型超市像如沃尔玛,家乐福等在线上冲击下的转变效果就显得差强人意。

1.1.2 传统零售行业近年发展分析

随着电子商务的增长式发展,传统的零售行业已经面临了巨大的挑战,许多传统的零售行业已经出现业绩的大幅度下滑,增长速度放缓已经成为传统零售行业的普遍现象,例如杭州解百集团股份有限公司发布的 2013 年的业绩报告,受互联网电商对传统百货业冲击的影响,净利润大幅下滑,2013 年实现营业收入 21 亿元,同比减少 2.19%,实现净利润 6296.39 万元,同比减少 18.63%,寻求其他的发展模式和传统零售的转型已经成为所有的零售商寻求发展的共识。在 2014 年 12 月 17 日,全球知名管理咨询公司贝恩公司在上海发布最新分析报告。据该报告预测,中国网上 B2C 零售销售额将以年均 25% 的速度保持增长,超出增长放缓的传统零售业三倍。通过 O2O 的模式,许多大型的零售商已经开始转型发展,例如 2014 年 2 月 11 日,北京王府井百货发布公告称,已与腾讯签署协议,结成战略合作伙伴关系。双方将基于微信公众平台商户功能、微信支付服务,开展具体合作。沃尔玛等也进行了官方网店的建设和运营,但大多数的效果并不明显。表 1-7 显示的是 2014 年前后几家大型零售商经营网店的运营状况。此外,华润集团的网上商城万家摩尔因经营方面问题停运;2013 年 5 月,永辉超市生鲜网站半边天上线月余就关停;卜蜂莲花线上业务也停产;2013 年 7 月,多渠道发展电商的德国最大零售批发超市集团麦德龙电商业务也出现萎缩,不仅网上商城的配送范围收窄至上海、江苏和浙江等城市,麦德龙天猫店也停止营业;2013 年 8 月,武汉中商的电商关闭。由此可见,传统零售商接触电商模式的尝试困难重重,效果也并不显著。由此,寻找传统零售商未来的发展模式势在必行。

实体店	电商名	状态(官网显示)
华润万家	万家摩尔	停业(网站进行重整)
永辉超市	半边天	下线
麦德龙电商业务	麦德龙	停止福州市配送 天猫旗舰店下线
武汉中商	中商购物网	下线
金福元	电商频道	停业(会员积分查询系统正在建设)
英国零售商 Argos	爱顾商城	关闭(已于 2013 年 3 月 18 日正式关闭)

表 1-7

1.2 研究内容

本设计旨在通过另一种消费模式的改变来寻求传统零售商的新发展,同时给人们带来生活上的更大便利。

1.2.1 冰箱智能化的可行性研究

智能冰箱早已在市场上流行,早在 2012 年底海信就推出了博纳智能 SMART 冰箱,在冰箱门上嵌有一台 10.1 寸屏幕,搭载 Android3.2 操作系统、内存高达 1G 的特制平板电脑,通过它以及专属的软件和感应器,用户不仅可以十分直观地对冰箱各个分区进行数字控温,更能借助定制软件,便捷的了解家中冰箱的食物种类、数量等信息。此外,它还可以自动提示食品的新鲜程度,根据用户的体质信息

提醒饮食的合理搭配,保证消费者饮食健康的最大化,是最佳的“健康管家”。在 2013 年三星公司推出 T9000,这款冰箱也内置了安卓系统,这款冰箱门配备了 10 英寸触摸屏,支持 WiFi 网络,自带笔记应用 Evernote,还可以向用户提供新闻、天气预报,甚至 Twitter 消息。

1.2.2 零售行业新营销模式的探索

尽管电子商务发展迅速,并且触及的商品门类繁多,但是线上购物需要至少一天的配送时间,多数配送时间是三到五天,因此生鲜食品很难实现网络上的购买;另一方面,生鲜食品由于其本身产品的特殊性,保鲜期较短,且无法承受巨大的碰撞,因此不适合以快递的方式进行配送。但是,生鲜食品的储备以及贩卖是大型超市的固有业务,另外,以沃尔玛的销售模式为例,沃尔玛是采取一个送货基地辐射周边几个零售超市的模式来实现商品的配送,以达到节约成本的目的,并且在条件允许的情况下,沃尔玛对于蔬菜瓜果类产品是采取自给自足,即在短距离的范围内进行种植,或者与果农合作供给等方式,由此成本可以大量地降低。因此如果直接将这些生鲜食品供应到用户的家庭,与供应到相应的超市相比,成本方面的差别不甚显著,并且超市的营业量可以有预见性的提高。

本设计的设计流程是在智能冰箱上连接无线网络的功能,设定一款冰箱专用的 app,通过 GPS 定位将辐射周边几公里以内的零售超市的官网进行连接,以实现用户可以在 app 内直接搜索浏览到对应超市的生鲜产品,实现在冰箱上下单,支付,确定配送信息,由此可以实现生鲜产品短距离,短时间的配送,甚至在几小时之内就可以收到货品。这样一方面实现了用户更为便利地购买生鲜食品的需要,另一方面对于传统的零售商来说,也可以以此为突破口寻求转型发展。(如图 1—8)

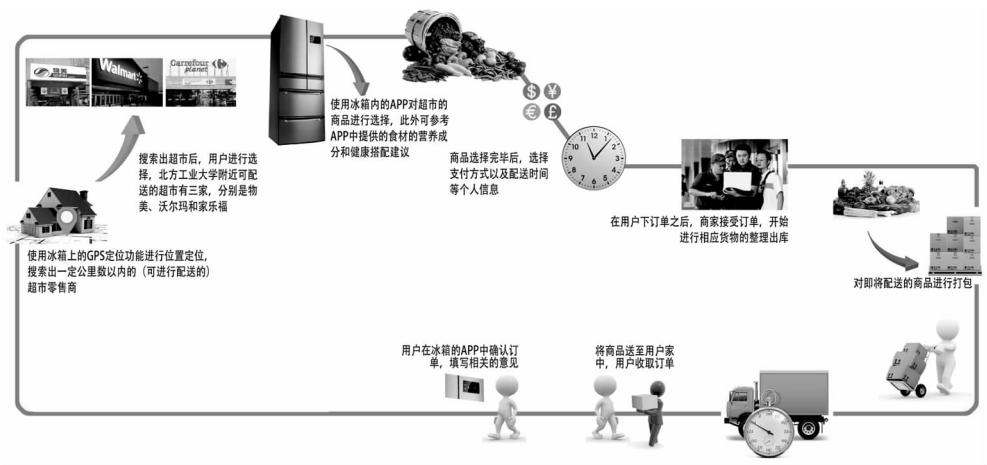


图 1—8

1.2.3 优势分析

1.2.3.1 对于传统零售行业的优势

超市百货等传统的零售商可以与电商企业争占市场份额
超市本身在生鲜食品上具有先天的优势,不需要付出额外的成本
因为超市的运送对象都是在一定公里数的辐射区内,因此不需要支付额外的物流成本。

1.2.3.2 对于消费者的优势

实现了瓜果蔬菜等生鲜食品的快捷室内购物,无需出门。
在当今社会压力下,人们长期处于亚健康的状态,食材的搭配和摄入的周期可以在 app 内设置定

时的规划和提醒,由此提高人们的生活质量和健康状况。

可以确切地了解某种食材的购入日期,以此来保证食用的食品有着绝对的新鲜度,并且可以通过记录来了解生活中哪些食品和营养的使用过量或者过少。

1.3 研究目的

本设计研究的最初目的是电子商务的发展迅速导致传统零售行业的尴尬境地问题较为显著,由此传统零售行业寻求新模式发展的需求迫在眉睫;而在人们的日常生活中,生鲜食品的网络购物无法实现,受到保质期,食品本身的抗压力等各方面的限制,生鲜食品很难真正实现网络上的长距离购物和运送,但是,鉴于当今社会压力的日益增大,上班族普遍没有闲暇时间在平时进行外出购物,因此生鲜食品的快捷购物也是当今快餐都市的人们亟需解决的问题。

1.4 研究意义

一方面是可以实现瓜果蔬菜等新鲜食品也能送货上门,并且在智能冰箱内添加科学的营养食谱,有利于人们在平时生活中更注意饮食健康,提升自身身体体质;

另一方面,在网络购物的冲击下,为超市等零售行业提供了新的可能性,谋求新的发展模式,以增强其与网络购物相竞争的实力。

2 智能冰箱的研究分析

2.1 冰箱发展现状分析

2.1.1 冰箱发展现状

我国家用电冰箱行业的发展始于1956年,在改革开放以后进入了飞速发展期,逐渐成为人们的日常生活电器。在二十一世纪以前,冰箱的发展趋势一般是较为注重制冷功能以及外观的设计方面;而自二十一世纪以来,冰箱开始偏向智能化的发展,温度的调节,制冷的变频等,与此同时,冰箱的空间也在不断增大,从最开始的老式上下双开门冰箱发展为法式冰箱、双开门冰箱、对门双开门冰箱、三门冰箱、多开门冰箱等。从空间的不断增大可以看出用户的市场需求:人们的生活水平不断的提高,导致用户对于冰箱的容量需求不断提高。根据中国家电协会信息部主任胡晓红的观点,随着家电产品普及程度的不断提高,冰箱的发展趋势将面向四个方面,第一是变频将成为普遍的冰箱配置,在未来的时间里,将完全取代定频冰箱的市场;第二是鉴于人们对于卫生的日益看重,对于冰箱的杀菌功能人们也会加大重视,因此对于冰箱的杀菌洁净能力也会有所提升;第三是未来十年的冰箱发展将会趋向于“高科技”“高附加值”,即对于冰箱的温度功能方面的需求以外,人们会对于冰箱有额外的功能要求和期待,此外,随着技术的发展和 innovation,冰箱的智能化方面人们也会有新的要求,对于加载 app 以实现冰箱的其他方面的功能也是智能冰箱发展的新趋势;第四是人们对于冰箱的健康化会有更高的要求,宋红燕描述说,“未来的健康冰箱应该是变频的、保湿的、杀菌的,更不会有老冰箱的异味。”

2.1.2 现有品牌近三年发展成果分析

近三年来,冰箱的发展日益智能化,从最开始的变频出现后,近三年智能冰箱基本都会将中间一层的空间设置为即可冷藏也可冷冻的空间,以此来为用户更为自由地使用冰箱的空间。此外,对于冷冻的霜等各方面的细节,在技术上也有着显著的提高。而在冰箱的智能方面,很多冰箱的品牌已经推出了实际的智能产品,而不再只是局限于概念产品的层面。

2013年4月,三星推出了一款名为T9000的智能冰箱,除了冰箱本身在外观上是四开门的设计外,最为吸引眼球的是这款冰箱加载了安卓的系统,这款冰箱门上配备了一块10英寸的触摸屏,支

持 WiFi 网络的连接,并且有自带的笔记应用程序 Evernote,此外,这款冰箱还可以向用户提供新闻、天气预报,Twitter 消息等。冰箱内的数据可以同步到手机或 iPad 等其他电子终端上,例如 Evernote 这款应用程序,里面的数据可以同步到 PC、Mac、iOS 设备或者任何同样使用这款应用的设备上。而其他应用像菜谱应用 Epicurious,用户可以选择一个菜谱,由此菜谱所需的食材可以生成一份购物清单,然后再检查清单上的食材哪些是冰箱中已有的,哪些是没有的需要购买的,这样用户就能更为清楚地知道所需购买的食材清单。用户通过对屏幕的触控,内置的其它应用程序能够让消费者与家人朋友分享照片,视频,和食谱等等。此外还可以内置一些其他的小工具,如日历和天气等多种应用程序。

海信也在 2012 年 11 月份推出了博纳 SMART 智能冰箱,这款冰箱可以搭载物联云服务,引入全新的食品管理功能,甚至能够及时提供食物的营养成分信息以及科学营养的饮食搭配推荐,自然也附带娱乐功能,由此让冰箱成为了一个厨房里的娱乐中心。这款冰箱是在冰箱门上嵌入了一个内存高达 1G 的特制平板电脑,10.1 寸的屏幕,搭载了 Android 3.2 的操作系统,通过这个屏幕用户可以十分直观地了解家中冰箱的食物种类以及数量等相关的信息,便于进行食品的管理。此外,它还可以自动提示食品的新鲜程度,根据用户的体质信息提醒饮食的合理搭配。这块平板通过内置的专门应用,主要是提供了两方面的管理,一个是对冰箱本身进行管理,例如选择功能模式、调节温度等,以前这些操作可能必须打开冰箱门才能操作。二是对冰箱内的食品进行管理,就是统计食品的数量、保险周期等信息,还可以通过它获取营养搭配建议或者食谱等。此外,除了管理冰箱和内部的食物之外,该屏幕还可以实现浏览网页、播放视频、显示天气预报等功能;其中内置了很多与饮食有关的应用程序,例如教做菜的菜谱、能制定家人健康食谱的软件等。另外,这款冰箱还内置了摄像头,通过这个装置,用户可以将食物的条码扫录进这个“平板电脑”中,条码中记录了食物的种类、生产日期和保质期等详细信息;同时内置软件会管理冰箱内所有食物的信息,例如提醒冰箱中哪些食物即将过期,以免造成浪费。海信的博纳智能冰箱,首次借助智能物联网技术手段实现冰箱的“食品管理”功能,将冰箱从信息“孤岛”借助物联云服务平台变成家庭智能终端。图 2-1 为此冰箱的界面设计。左图为这款 app 应用程序的主界面,可以看到里面装载了不同的应用程序,已实现了不同的功能。右面两个图是同步到手机的用户信息,登陆个人的账号后,可以看到购入的清单查询以及食品的日子查询,由此来确定食物的保鲜期,以防食物的变质和浪费。



图 2-1

另外, LG 也在 2013 年推出了一款名为 LFX-31995 智能冰箱, 内置 NFC 连接移动设备, 同时能够连接 WiFi 网络。从主界面来看, 不确定是否使用的安卓内核, 不过据说可以和 Android 和 iOS 设备兼容, 以实现信息与手机的同步。该设备一方面可以直接通过这块屏幕进行冰箱的温度设置等, 另外通过在手机上安装 APP, 同样能够实现对冰箱的设置。此外, 除了对冰箱的食物相关方面的信息, 这款冰箱还内置了天气预报、备忘录、照片浏览和谷歌日历等应用程序, 照片是通过无线方式从手机获取。

2.1.3 冰箱使用的功能性分析

此次设计的智能冰箱需有达到的功能主要有以下几个方面:

1. 购买食材的需求。一方面是为了满足人们不出门就可以购买新鲜食材的需要, 使得日常的生活更为方便。另一方面是为了寻求一种新的商业模式, 来帮助传统零售业面对业绩下滑的窘境, 以更好地应对电子商务的迅猛发展所带来的冲击。

2. 鉴于当今社会压力较大, 人们在长时间的疲劳工作和不正规的作息状态下, 很多人的身体处于亚健康的状态, 因此此次的智能冰箱需要提供食材的营养成分以及营养搭配摄入的推荐, 以改善和提高人们的身体健康状态。

3. 食谱的提供和推荐。鉴于很多年轻人并不擅长烹饪, 对于所烹饪的食物没有把握, 因此在应用程序内添加食谱的信息, 可以让不善烹饪的人以此为参考, 作出较为喜欢的美食。

4. 交流的功能。用户可以在这里分享各种经验、心得、食物方面的最新资讯, 或者是其他的内容, 使得人们更为重视我们所摄入的食物, 了解怎样更为健康地饮食。

3 交互设计的调查研究

3.1 交互设计概念

3.1.1 交互设计定义

“交互”和“互动”最早可以追溯到历史上早期人类在狩猎、捕鱼、种植活动中的人与工具之间的关系。在《辞源》中, 对“互”的解释是: 交错、互相、彼此之意。“交互”或“互动”一词在英文中出现较早, 为 Interact, 意为相互作用、互相影响、相制和交互感应。“交互”或“互动”中文一词是在现代语言环境下, 特别是计算机信息技术广泛影响社会以后产生的新汉语词汇, 在中文中原属社会学术语, 指人与人之间的相互作用, 分为感官互动、情绪互动、理智互动等, 指共同参与、互相影响、互相作用。随着计算机等数字媒体的发展, “互动”在这一领域里特指人机之间的相互影响和作用。计算机交互技术的出现, 是人与人之间情感的互动开始转移为人与计算机之间情感的交互。简而言之, “互动”可理解为人与人或人与物相互作用之后, 给人感官或者心理上产生某种感受的过程。^[2] 维基百科的解释是交互设计在于定义人造物的行为方式(the “interaction”, 即人工制品在特定场景下的反应方式)相关的界面。

交互设计师首先进行用户研究相关领域, 以及潜在用户, 设计人造物的行为, 并从有用性, 可用性和情感因素(usefulness, usability and emotional)等方面来评估设计质量。

人从出生开始就利用感官、想象、情感和知识与周围的产品和环境进行某种形式的对话。而交互设计就是创建新的用户体验, 其目的是增强和扩充人们工作、通信及交互的方式或交互空间。正如交互设计师 Terry. R. Schussle 曾经在 Mac World San Francisco' 1988 年的会议指出的那样: 互动既不是动画, 也不是视频, 它是“用户控制和活动体验”。交互设计的早期著名理论家之一 Terry Winograd 在 1997 年曾经把交互设计描述为“是人类交流和交互空间的设计”。同样, 毕业于卡耐基·梅隆大学

的交互设计师 Dan Saffer 也认为:“交互设计是围绕人的:人是如何通过他们使用的产品、服务连接其他人”,Dan Saffer 还绘制了一张学科关系图,以描述交互设计与其他相关学科的关系。(如图 3-1)他指出交互设计其实就是沟通的设计,不同于说话、歌唱或表演等,交互设计实际需要的是产品或服务作为媒介,但本质还是人与人之间的交流。而这里所指的“交流和交互空间”或“产品与服务”正是媒介或媒体。

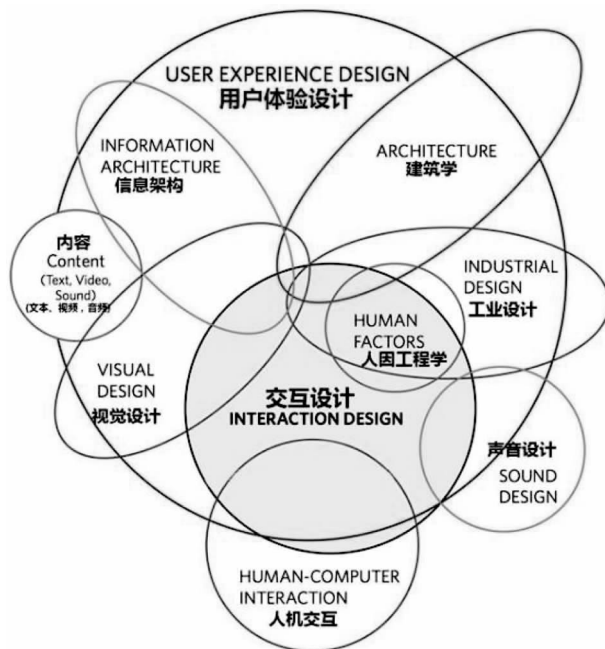


图 3-1

如图所示:交互设计是一个跨学科的交叉研究领域,交叉领域包括信息架构设计(Information Architecture)、视觉传达设计(Visual Communication Design)、工业设计、认知心理学和人因工程(Human Factors, 或人机工程学)、用户体验(UX)设计、人机界面或人机交互设计(Human-Computer Interaction, HCI)等。

3.1.2 交互设计发展脉络

交互设计虽然是近年才出现的学科门类,但交互设计的思想其实古已有之,例如中国历史上以烽火台或狼烟为信号传递远距离的消息和信号,因纽特人以石斧等工具修建家冢纪念碑作为地界标志等都是交互设计思想的体现。1837年, Samuel Morse 设计出了莫尔斯密码,将简单的电子脉冲转化为语言分类信息,从而实现远距离传送信息的目的。

自 20 世纪 30 年代以来,交互设计经历了人机界面到人机交互的发展历史,其主要包括三个阶段:第一个阶段是 1959 — 1970 年。

第二个阶段是奠基期,时间是在 1970 — 1979 年间。

第三个阶段是在 1980 — 1995 年间。20 世纪 80 年代中期以微软为代表开始了“个人电脑的时代”,此时,计算机终于进入了 GUI(图形交互界面)的时代,个人电脑开始成为人们办公、生活娱乐的新形式,软件的发展,电脑技术功能的复杂化以及大量非专业用户的需求成为了交互设计发展的最初的动力。与此同时,人机工程学、工业设计、人机界面设计和用户体验研究等领域也取得了显著的成

就,为交互设计的发展提供了助力,特别是基于人类信息加工理论的认知心理学的研究成果,如学习、记忆、疲劳、注意、情感和视觉的生理心理机制的研究为交互设计提供了重要依据。20 世纪 80 年代中期,硅谷的工业设计师比尔·莫格里奇(Bill Moggridge)在一次设计会议上提出了交互设计的思想,随后他把这个新设计领域称为更准确的“交互设计”,它关注如何通过了解人们的潜在需求、行为和期望来提供设计的新方向(包括产品、服务、空间、媒体和基于软件的交互)。交互设计是随着二十世纪九十年代互联网的兴起而发展起来的,该时期出现的电子邮箱、网络游戏等一系列针对个人服务的需求极大地推动了交互设计的发展。

3.1.3 交互设计发展特征

唐纳德·诺曼在《情感化设计》中将设计分为三个层次:本能水平的设计、行为水平的设计、反思水平的设计。不同的设计层次有着不同的设计目标,由此产生了不同的交互设计方向。^[8]

用户行为	用户目标	交互设计方向
本能层次:感觉的、安全的、直觉的、瞬时的反应	体验目标:简单的、个人的乐趣,酷时尚、成就体验	情感设计:吸引人的、感兴趣的、美观的、愉悦的、外观的
行为层次:理性的、认知的、直觉的、瞬时的反应	任务目标:完成作业、完成作品、完成计划	行为设计:使用的乐趣和效率、可用性、用户体验
反思层次:文化教育、意志、思考、解释、理解、推理	人生目标:美好生活、成就自我、功利目标、价值观	反思设计:自我形象、满意、品牌、记忆、长久行为

3.2 交互设计的两个维度

随着技术的发展,电子产品的搭载终端也日益多样化,因此交互设计使用的领域也随之扩大。由于电子终端的变化的多样性,在我看来,交互设计的发展趋向于两个维度的发展。

第一个维度是界面本身设计的形式与人之间的体验交互,用户使用的满意度,在接触时的被吸引度以及接受界面时难易程度,都会对用户的使用造成感官上的影响,以此来判断该交互设计是否设计合理,存在的相应的缺陷以及如何优化整理。交互设计本身在很大的程度上是人机之间的交互,因此“人”与“机”之间的联系就是交互设计的第一个维度。

第二个维度是界面设计与电子终端之间的交互,基于电子终端的多样性,界面本身也会随之产生相应的变化。例如,同样一款应用程序的界面,搭载在不同的电子终端上,界面的设计也会有所不同,如图 3-2 所示,是腾讯视频在不同媒体终端的 APP 界面设计。



图 3-2

在搭载电脑终端的界面设计里,由于搭载终端的界面尺寸较大,有足够的界面空间来容纳所需的信息,因此在界面的顶端除设置必须的导航条外,将推荐的视频放满整个屏幕,以达到夺人眼球的效果。搜索栏,导航栏以及用户的登陆等信息在首页面中一目了然。浏览方式自上而下,采用滑轮的形式,今日热门、原创精选、电视剧、综艺、电影以上下的布局进行排列,至于一层页面层级中。由于是鼠标滑动操作,因此导航条的次级信息分类采用浮动窗口的形式(如图 3-3),并且顶栏的推荐页面的切换也采取以鼠标滑动切换的方式来进行阅览。

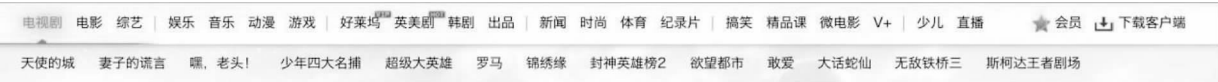


图 3-3

而在搭载 IPAD 终端的界面设计里,则对界面进行了信息的分类,将电脑终端中上下排列的项目栏以分类标签页的方式进行了分布,在首页中,进行了两种方式的分类,分别置于页面的最上方以及页面的最下方,由于 IPAD 终端的操作方式以手指触碰屏幕和滑动屏幕为主,因此切换的方式是采取单击的方式和滑动的方式来进行页面的切换。上方分类栏中的今日推荐、同步剧场、精彩大片等可通过手指左右滑动的方式来进行切换,也可通过单击项目块的方式进行切换;而下方的分类栏则通过单击的方式来进行内容的切换。搜索、个人中心等内容进行了信息的归纳,如以个人中心为例,用户的登陆,观看历史,收藏,缓存信息等都在个人中心中以子菜单栏的形式进行布局,采取的是左右分栏的样式。而在电脑界面里,这些信息则出现在首页面的顶栏右上角,用户的登陆,观看历史以及收藏、上传信息等都出现在第一界面。(如图 3-4 所示)



3-4a IPAD 个人中心首页



3-4b 电脑终端个人中心信息位置

图 3-4

最后在安卓手机终端的界面设计里,由于界面本身的局限性,对于信息的归类就更为细致,信息的层级划分更为复杂,由图九可看,在页面的顶端时导航条的分布,由于手机终端的使用习惯以点按和滑动为主,因此在界面受限而又不影响整体阅读的情况下,导航条的信息采取左右滑动,单行分布的形式进行布局。下方主页面则是采取上下滑动的方式进行布局,所有的信息以上下分栏的形式进行分布。此外,由于屏幕尺寸本身的局限性,相对于其他两种形式的终端来说,对于信息的内容作了相应的删减,比如在其他两种终端中的原创精选在手机终端中就被删除。

此外,依据电子终端本身的尺寸问题,界面的切换方式也进行的相应的改变。例如 IPAD 的界面是长多于宽的尺寸,因此界面之间的切换多以横向切换为主;而手机的尺寸则是宽多于长,因此界面的切换则以上下切换为主。这是依据终端本身的尺寸限制条件所做出的改变。这是交互设计第二个维度的交互,是根据终端的不同,来对界面设计做出不同的设计布局。

而在第二维度还有另一个方面的考虑,就是界面终端的形式所造成的区别。随着技术的更新发展,对于传统的产品上搭载屏幕进行智能操作的趋势已经日益明显,像如现实世界中已经实现的车载屏幕,iwatch 屏幕,谷歌此前所出的 Google Project Glass 等等,都是传统生活产品搭载智能屏幕的成果,例如在车载屏幕中,由于屏幕外有外置的按键,因此在屏幕的切换上就多以按键的设置为主,如果按键设置中有“左”和“右”的指向按键,那么在屏幕的选择中选项的分布也多是横向为主,来进行左右的切换。而在 Google Project Glass 的界面设计中,由于是贴近眼睛进行界面的展示,为了达到更好的效果,则多以浮动窗口层叠的方式进行展示,并且所有的内容也都尽量在首页面中完全展现,至今进行分层级的选择。当前以传统产品搭载屏幕的方式较为少见,且用户群体为少数,因此无法进行更为细致的分析,也并没有在市场中形成一个独立的门类。但在人们对于未来的虚拟产品展望中,交互设计所搭载的产品终端却是五花八门,桌面、镜子、墙面等,凡是光滑平整,色彩干净的平台都可以通过光学反射原理来予以实现。因此终端的不同,导致界面交互的方式更为多样,由此对于界面设计的要求也就会有所不同,这也是基于终端的不同而对交互设计本身所产生的影响。

3.3 交互设计感性设计思维与理性设计原则

3.3.1 交互设计感性设计思维

交互设计是基于人的情感式体验,来作为设计的目的。而另一方面,由于交互设计真正进行发展的时间只有近十几年,因此交互设计本身并没有什么约定俗成的设计规则和准则。而更多考虑的是人

们的心理情感,使用时的便易度,以及人们的切身感受。所谓感性设计在我看来就是依据大多数人的使用习惯或是思维定式所导致的习惯进行分析,由此来进行分析,设计出相应的浏览方式,这样人们会依据惯有的思维方式来进行页面的浏览,就会达到自然而然的效果,没有任何不舒服或使人感到别扭的操作,就达到了用户满意度的考虑。

第一个是人们的阅读习惯,在非阅读性或新闻类的内容里,人们通常不习惯进行大量的阅读,因此在 app 设计中,一般不会出现大段的需要阅读的内容,多是以标题的形式出现,属于眼过即阅完的长度,此外即使在电脑或是 ipad 等屏幕较大的条件下,一般也会采取分栏的形式对内容加以编辑,不会出现整段的区域来进行内容的布局,这样会容易让浏览者产生疲劳的感觉,长时间的浏览会加大浏览的压力,由此对于浏览的舒适度会有所降低。即便是在新闻类的界面设计里,也会注意阅读的舒适性,以今日头条和网易新闻的 iPad 应用程序为例(如图十三所示),整体的界面都是分为三栏,文字大约都在十个字到二十个字,因此用户可以轻松地浏览整体界面,再加上图片的搭配,整个界面就会更具有吸引力,引导用户进行阅读。即使在单个新闻的页面报道中,新闻内容每行也会少于二十个字,并且以分栏的形式对 iPad 的屏幕进行划分,以此来达到阅读更为顺畅的目的。如图十三中最右侧的图片所示,该图片为网易新闻 iPad 的应用程序界面,界面将左栏设计成单个新闻的引导栏,将右栏设计为新闻的阅读栏,并且通过添加图片的方式对新闻的阅读长度进行压缩,以达到最适宜的浏览长度,使用户获得最为舒适的阅读体验。这是依据人们的阅读习惯所作出的交互界面的设计。



图 3-5

第二个是人们的浏览习惯。人们在当今社会的浏览习惯一般是由左往右,从上至下,因此在交互界面的设计中一般也是以此为设计依据。大多数的界面首页面的导航栏如果设置在页面的顶端或是页面的最底端,那么多是以左往右的阅读顺序,如果是将导航条设置在页面的左侧,那么多数是从上至下的排列。这样符合大多数人的阅读浏览习惯,因此在页面设计中也大多以此设计,以达到最大的舒适度。

第三个是基于操作的习惯。有许多的页面布局以及操作习惯是约定俗成地应用在不同的电子终端以及不同的应用程序中。比如在界面的左边栏如果是有右志向的浮动箭头,那么这可以进行向右的滑动,出来的界面一般是用户的个人设置和用户信息面,或者是导航栏。信息的分栏区域一般选择的方式是单击,长按多为对其进行编辑,页面的左右滑动,上下滑动等都已经成为一种约定俗成的界面设计的习惯,因此用户对于这种操作较为习惯,在舒适度上已经自然而然,如果进行更改,那么会引起用户的反感,在使用舒适度和满意度上也会相应地有所降低。

第四个是针对特殊用户群体的考量。比如一个关注老年人健康的应用程序,那么用户的主题应该是老年人人居多,那么在字体的设置上就要相比一般的应用程序更为清晰、明了、易读。如果是针对儿童

群体的一个应用程序,那么在色彩上就要更为亮丽,更为童真,这样能真正吸引到儿童的实现,使得他们更愿意使用,由此来提高相应的用户满意度。以苏宁易购和美丽说为例,两个门户网站同为购物的门类,但是针对的消费群体却有所不同:苏宁易购的主题产品是家用电器,购买的用户较为繁杂,但是由于电器本身的价格较高,因此不属于冲动消费的商品,因此苏宁易购的官方网站的主色调为蓝色;而美丽说针对的用户群体则为女性,更为针对性地说是相对年轻的女性,这个用户群体消费观并不足够地理性,容易冲动消费,并且容易受到图片或其他方面的影响,因此该网站所用的主题色调是年轻女性比较偏爱的粉红色。(如图 3-6 所示)



3-6a



3-6b

图 3-6

3.3.2 交互设计的理性设计原则

“好的产品一定是建立在对用户需求的深刻理解上。”交互设计之“交互”一词就已经说明了用户的重要性,而交互设计的设计原则,就是为了保障“人”与“机”之间的交流的流畅性,尽量减少障碍度。因此交互设计有以下几个设计原则:

1. 置界面于用户的控制之下

有很多网站出于其他的考虑,例如商业利益、营销等层面的考虑,会实现帮助用户预设一些决定,但是这样所引导出的界面有时会为用户所不愿意或者反感的事情。这样的举动就严重干扰了用户的操作进度和用户目标的完成,给用户带来了不好的用户体验,阻碍了“人”与“机”之间沟通的流畅度。例如现在在进入一些网站的应用程序时,会出现三秒的广告界面,比如乐视网、有道、ireader 等,这样在用户打开界面的初始时间里就被迫等待了三秒的时间,并且无法进行任何想要的操作,因此在进入应用程序时就有了一种程度的反感和负面印象,这样对用户的使用满意度有很大的影响。

2. 学会换位思考,以目标用户的角度思考问题

设计师在设计产品时要能够做到换位思考,体会用户的立场和感受,并站在用户的角度思考和处理问题,把自己置身于相关的用户场景中,理解用户的行为特点和行为差异。之后,从目标用户的角度来进行功能、使用情感等各方面的考虑,从而设计出应对相应问题的成果。例如给商务人士使用的新闻类应用程序,应考虑到大多数商务人士的思维模式定式,商务人士由于其工作的原因,大多数人更偏向理性和简洁的思考模式,讲究直接明了,而并不注重繁琐却没有意义的纹饰图案,对他们来说这只会带来视线干扰的效果;但从另一个角度出发,如果是一款设计给青少年以及儿童的应用程序,那除了本身功能上的满足外,还要考虑到界面上色彩、排版等是否足够具有视觉上的吸引力,这种吸引力本身就

会影响目标用户对于此款应用程序的使用频率。另一方面,如果是针对某一特殊人群使用的应用程序设计,那么就要考虑该特殊人群的生活习惯。比如设计一款应用程序,用于帮助孤独症、自闭症人群更好地、更为积极地了解社会,那么就要注意信息内容的摘取,注意遣词用句,包括自闭症人因为并不善于接触新事物,没有太大好奇心的特点,因此在应用程序的设计上就要注意不能有太多的子栏目以及太过繁杂的链接,否则会引起他们的反感和恐惧,这样就达不到最为初始的目的。

3. 最大程度地注意信息的给予性,而不要希冀于用户的记忆能力

用户在使用一款应用程序时,注重的是应用程序本身的流畅感和体验感,用户在浏览内容时针对其感兴趣的方面进行浏览,在切换下一页面后,往往希望看到的是前一页面的延续深入或是另外一个层面的内容,如果用户在浏览后页面时需要对前一页面进行回忆,那么记忆力不好或是无心记忆的人就要不断切换页面,以完成对整个内容的浏览和理解,这样会无形中造成用户浏览的障碍性和不流畅性,这样会严重影响用户使用时的情感体验,造成不良的后果。

4. 设计的整个页面风格要保持一定的一致性

每款应用程序在设计之前都有内容的梳理,用户人群的定位,功能上的要求以及所要体现的特有的风格,比如美丽说的应用程序,定位的用户群体为 15—37 岁左右的少女,主要的商品为衣品配饰以及护肤品,因此在应用程序的色彩选取上选择了少女系的粉色,以便更为贴合目标用户的审美。不论是在浏览商品界面,还是在购物界面,或是商品的详细介绍界面,美丽说的整体风格都保持了粉底白字或是白底粉字的风格。所有的商品图片为了更为贴合整体风格,图片本身也多以白底为主,图片整体偏轻色系,由此使得整体的风格更为统一,风格明确。应用程序内的 logo,图标,文字等也在色彩上实现了统一,保持了整个界面风格的一致性,由此整个应用程序就给了用户一种更为直观的感受。

5. 视觉效果占主导地位

Stephen P. Anderson 说:“视觉影响行为,也影响体验。”在人的味觉、触觉、嗅觉、听觉、视觉这五类生理感觉中,视觉的感官是最迅速的,信息图专家 David McCandless 认为,人类的大部分的大脑资源都花在了视觉上。由此在应用程序中,虽然我们会收到听觉或是其他方面的影响,但基于这样一个特殊的平台,视觉方面的感官体验才是我们最应该给予重视的。更深层次的方面来看,Stephen P. Anderson 指出,视觉影响的不只是体验,也会影响用户的行为。在应用程序中除了某些特定的、约定俗成的固件的图表展示外,整个应用程序的使用方式和视觉顺序还是依靠设计应用程序时设计师预定好的潜在顺序来进行浏览,而用户会依据潜在顺序进行浏览的原因就在于设计师在应用程序中提供了一定程度上的视觉信息来进行用户视觉的引导,由此实现引导的效果。

4 智能冰箱的界面设计分析

4.1 界面的尺寸条件分析

界面交互的载体的尺寸影响着界面交互的设计内容和具体的交互方式,比如同样一款应用程序,搭载在电脑终端和搭载在 iPad 终端的设计方式就明显不同,原因是终端尺寸的大小就影响着界面所能容纳的信息量,因此 iPad 的终端相对电脑终端的界面来说,就要进行信息的压缩和整理,分的子层级相对来说也要比电脑终端更为复杂,页面的布局由于链接的增多也就要考虑得更为精细细致。

此次智能冰箱的尺寸为 3:2 的尺寸比例,而内部屏幕的尺寸比例为 16:9,屏幕的大小大约 170mm * 100mm 的尺寸,横屏。现在普遍使用的系统为安卓系统和 IOS 系统,由于按键方式的不同会导致用户有着不同的使用体验和使用习惯,例如 android 的返回键很多时候代替了 iphone 界面的 back 键;但

是即便在按键设置上有着较为明显的差异,在页面布局、跳转、层级结构、文案、提示语等方面还是要保持一致的,另外在细节方面要贴合平台特性,以及界面设计的原则和风格。

因此基于本次屏幕尺寸的条件,界面的设计如图 4-1 所示,中间的主体部分是商品的图片信息以及较为简略的商品介绍,下方的点状物显示的是页面的信息,当前显示的是总共有 10 个界面,用户通过手触的方式左右滑动,来进行页面的切换,浏览不同的商品信息;顶栏显示的是时间,网络状态等基本固件信息;下面紫色区域是整个应用程序的主菜单栏,由于该应用程序没有任何外部可操作的按钮,因此用户可以通过下面的菜单栏来进行更为直接的页面切换,选择相应的操作模式。

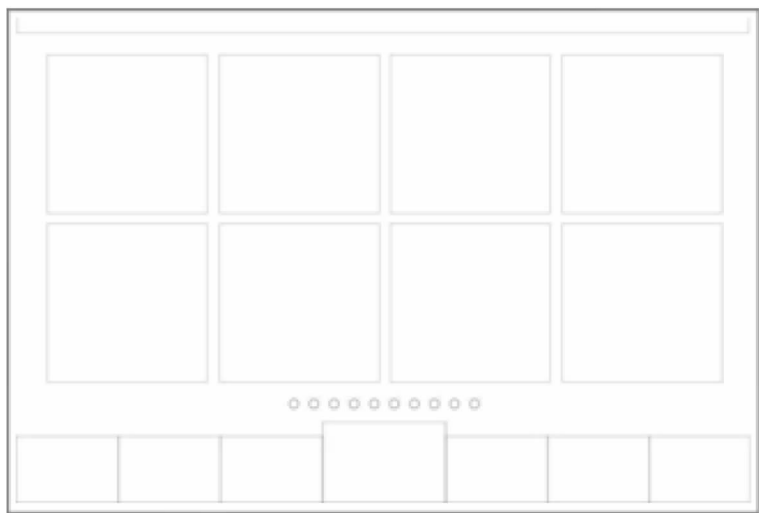


图 4-1

4.2 用户的定位分析

此次设计的应用程序多用于生鲜食品的采买和烹饪、营养等方面的引导,使用的用户更多的是善于烹饪或是勤于家务的女性,年龄一般在二十岁以上。在设计之前的用户调查中,针对不同年龄阶段的用户进行了问卷调查,各人群年龄段划分分别是 15—20 岁,21—30 岁,31—40 岁,41—50 岁四个组别。调查的结果显示 15—20 岁和 21—30 岁的两个组别对这种新的购物模式认可度较高,并且在意见中普遍希望实现手机的同步,31—40 岁的用户其次,而 41—50 岁的用户组则对新的购物模式兴趣不高。

此外,基于智能冰箱的成本问题考虑,搭载智能屏幕的冰箱售价应该会在中上的价位,因此大多数的年轻人并不具备购买的经济条件,因此定位的用户群体应在 30 岁左右的女性和热爱烹饪或是熟于烹饪的少数男性。

4.3 界面的色彩设计分析

此应用程序是搭载在冰箱上,主要的功能是为了让人们对于瓜果和生鲜食品的购买行为变得更为便捷,以及改善人们的饮食习惯和身体健康。基于此目的,在颜色的方面首先定位的是偏感性的暖色,例如红、黄、橙色等,而不是偏感性的冷色,像如蓝、青、绿等;其次,由于现代人观念的改善,烹饪饮食已不再是专属于女性的标签,男性的烹饪比例越来越高,因此在暖色调的选区里,首先放弃了一些专属于女性的颜色,例如粉色;再次,考虑到很多的家居设计将冰箱放在了厨房区域,而厨房忌火,不宜选用太过热烈张扬的颜色,因此将红色也排除在外。在三原色的色彩体系中,暖色就只有橙色和黄色可用,但

由于黄色的色调不甚明显,经过自然光的反射后会不宜辨清,不适宜作为应用程序设计中的单一主色调,因此最终选用橙色作为本次应用程序设计上的主色调。

橙色是暖色系中最温暖的一个颜色,是三原色基本色系中红色和黄色的混合色,代表着欢快活泼的特性,具有明亮、华丽、健康、兴奋、温暖、欢乐等性格特点;此外,橙色使人联想到硕果累累、丰收的金色秋天,是满足和喜悦的颜色,因此,橙色作为一款以食物为主要内容的应用程序的主色调在其色彩的本质精神上也是相当契合的。

5 新零售模式的研究与可行性分析

5.1 沃尔玛模式探析与适应新模式的可行性分析

沃尔玛作为零售行业的翘楚,其营业额一直在中国市场处于领先的地位,因此以沃尔玛的销售模式对此次新零售模式的可行性进行分析。

沃尔玛的经营模式中最为亮眼的就是他的低物流成本,由此使得他的整体成本要低于行业内的整体水平,这样在商品定价低于他人的情况下可以实现最大程度的盈利。以下是针对沃尔玛的经营模式来对新零售模式的可行性进行的逐条分析:

1. 沃尔玛的选址

沃尔玛公司的选址条件一向以当地居民为首要条件,一般选在整个城市较为繁华或者是该地址的居民居住数量到达几十万以上,由此来保证该地域内首先已经有一定的消费能力。并且沃尔玛对于该地居民的消费能力也有一定的要求和调查,一般此地的居民的收入水平应该要在年收入三万美元以上,即当地域的居民有足够的购买能力来进行生活的需要的补给。因此沃尔玛的选址有一个先天的优势,就是首先保证了消费客户群体的一个基本量级。对于新的零售模式来说,由于本身针对成本的问题考虑,该零售模式在运行初期不建议搭载外来的物流公司,因为这样会使成本迅速提高,增加营运的风险;并且此零售模式的优势就在于快,用户在下订单后,超市可以在两个小时之内送货到家,完成用户对于商品的采购,这就要求用户与超市之间的距离不能太远,而沃尔玛的选址条件就因此契合了这种新的零售模式。对于一个基量在 50 万以上的用户的居民群体来说,新零售模式已经有了一定数量的潜在用户来保证其运行的可行性。

2. 沃尔玛的商品结构

沃尔玛的商品种类齐全,涉及到人们生活的方方面面,几乎是应有尽有。基于沃尔玛本身对用户群体的定位,因此沃尔玛的商品主要以中档的商品为主,高档的和低档的商品适当兼顾。在某些商品方面,例如生鲜蔬菜,如果沃尔玛可以在配送中心附近寻找到适宜的土地,那么沃尔玛通常会销售自己出产的生鲜食品,这种情况在蔬菜方面的情况较多,这样可以很大程度上节约了成本,也保有了食品本身的新鲜度和质量。从新的零售模式方面来考虑,附近的用户基本都满足沃尔玛最开始选址时对于附近居民的消费定位,因此附近的用户的购买能力正是贴合了沃尔玛的商品定位。

3. 沃尔玛的物流战略

沃尔玛为了保持它的低成本条件,一般会采取建设物流配送中心的方式,一个配送中心大约辐射周围 320 公里以内的商圈,大约有一百多家沃尔玛的超市门店,这样就使得这一百多家门店的商品供给可以灵活调节,并且商品的物流成本降到最低。针对新的零售模式来说,沃尔玛可以不需支出任何额外的成本,在商品订单下达到沃尔玛的电子系统内后,沃尔玛可以从配送中心建立专门的送货系统,来进行配送,这样就不需要将商品配送到卖场以后再进行用户家庭地址的配送,可以直接将商品从物

流中心配送到用户的家庭住址。

5.2 新销售模式的具体流程

5.2.1 人员需求和安排

从人力成本上来说,新的销售模式不需要增加太多的人员成本,以沃尔玛为例,它本身有大量的超市配送人员和营业人员,如果人们习惯于这种新的零售模式,那么相对的,超市的在店人员就会相应地根据消费者流量的减少而可以适当地减少,这样即便在一开始需要额外的人员物流方面的增加成本,这些成本对于一个大型的零售超市来说也不会太多,风险也不会太大。

5.2.2 流程的示意

整个流程示意如下:当人们购入冰箱或是搬到新的地方后,就可以用冰箱内的 GPS 系统进行定位,扫描附近的大型超市,扫描的距离以该超市预设的配送公里数作为依据,当扫描出所有的附近可以进行配送的大型零售超市后,用户就可以在应用程序内选择商品。

以北北京市石景山区北方工业大学内的家属楼为例,图 5-1 是北京市石景山区的卫星俯视图,阴影部分显示的是北方工业大学的地理位置,在左下角放大的图片中,阴影部分显示的是北方工业大学内家属楼的住宅位置。通过 GPS 系统定位以后,扫描出在北方工业大学附近有三家大型的零售超市,分别是物美,家乐福和沃尔玛,这说明这三家的地理位置都在可以配送的范围内。在确定了可以配送的超市后,应用程序内出现的商品就会是这三家零售超市可以出售的商品信息。用户在购买商品前,在应用程序内有营养搭配的参考信息以及各个商品中所含有的营养成分信息,用户可以根据这些信息来有选择地购买所需的商品,使得用户所购买的商品可以更有针对性地实现对健康饮食的要求,改善人们的健康状况。在商品选择完毕后。

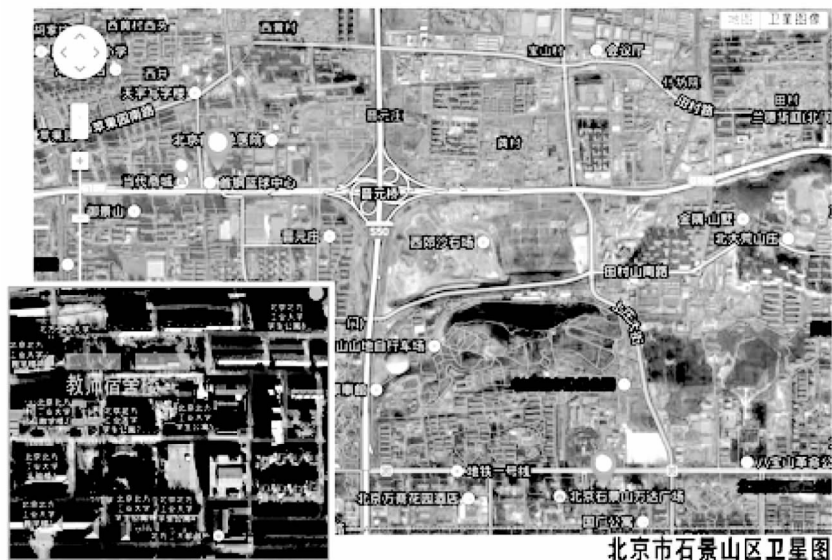


图 5-1

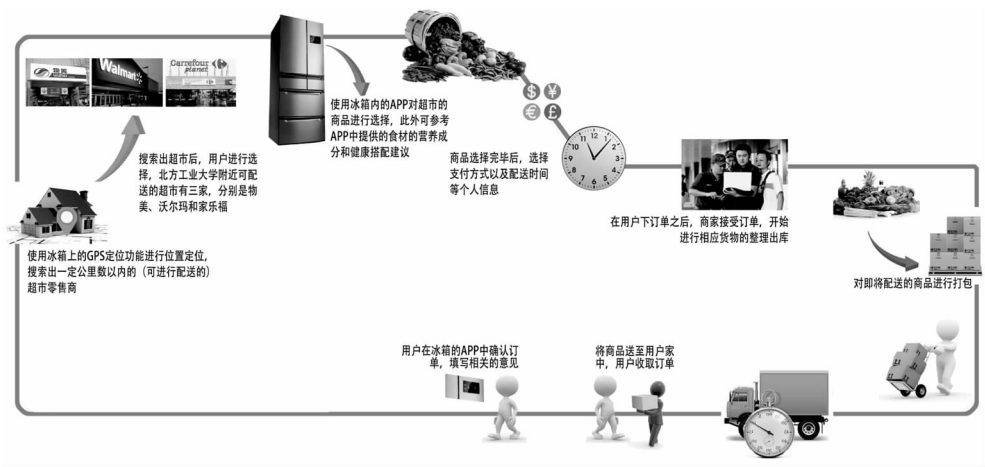


图 5—2

用户就可以确认订单,选择配送的时间和个人的联系方式等相应的配送信息,在用户下单后,超市可以第一时间收到订单内的商品细则,进行相应商品的拣选整理,然后出库,之后对商品进行统一的打包,就可以进行商品的配送。商品送到用户家中以后,用户签收订单,之后可以在应用程序内进行订单的确认并填写相关的意见和建议。由此完成整个的购买流程。

5.2.3 新模式的意义和优势分析

新的零售模式有两方面的意义,从传统的零售商来说,可以寻求一种新的商业模式,来应对电子商务的蓬勃发展对传统零售商的冲击,改变近几年来销售业绩大幅度下滑的尴尬状态,重新抢占市场份额,实现更大的利益化;从用户居民的角度来说,人们可以不用出门就实现生鲜食品的采购,一方面节省了大量的时间,在面对工作压力的情况下,可以不用分神来用额外的时间去满足生活上基本生理的需要,另一方面,用户依据营养参考的信息进行食品的购买,可以使人们越来越重视食品的营养成分和健康度,使得人们对于自身的身体健康状况更为了解和关注,从而改变近年来年轻人身体状况长期处于亚健康的尴尬局面。

此外新零售模式的运营有一系列的优势条件,一方面是市场需求方面,基于这种零售模式的便利性,用户在时间不足,生活压力大的状态下更愿意选择不出门就能购物的方式,这样就有大量的潜在用户群体,即有相当乐观的潜在市场规模。另一方面,由于传统的零售市场在电子商务下冲击巨大,并且做了一系列的尝试,包括开启 B2C(Business to customers)模式的网上商店,与已有的网络电子商务平台进行合作等,但是效果甚微,成果也不太理想。从传统的经营模式转型,与纯粹的电子商务平台进行竞争,从形式上来说进入了相对陌生的领域,与已经有大量经验的对手进行竞争,这种竞争本身就是存在缺陷,以己之短攻其所长,没有什么竞争优势。而这种新的销售模式却利用了传统的零售商本身所具有的传统优势,以另辟蹊径的方式与电子商务平台征战市场份额,进行竞争,这样才有了竞争的优势和条件。

6 智能冰箱应用程序界面的设计内容和交互详述

智能冰箱的应用程序依据实际的需要添加了以下几个方面的内容 ,依次分别是个人用户中心界面,搜索界面,食谱的参考,应用程序主界面,收藏界面,购物车以及交流分享区域。

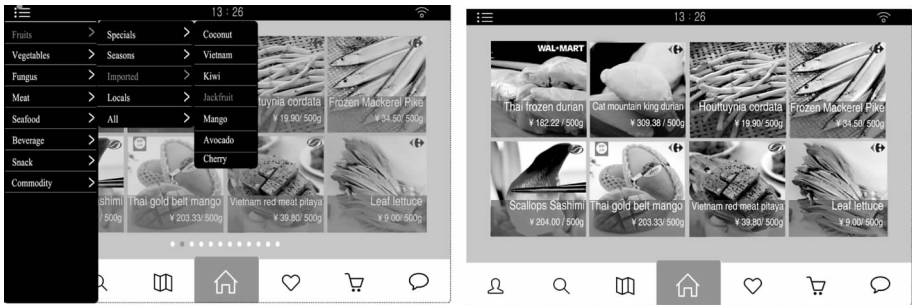
6.1 主界面

主界面是进入应用程序后的第一个界面,界面的中心区域一般是当天零售商推荐的特价商品或是较为优质的食材。根据屏幕的尺寸大小和其操作的特点,页面之间的切换方式是以左右滑动来进行切换,下方的点数显示的是界面的总体页数,用户可以点击自己感兴趣的商品进入该商品的详细介绍界面,来选择购买或是进行收藏。商品的界面将商品的图片进行了最大程度上的放大,一边用户更为方便地对商品进行浏览(如图 6-1),用户的左上角又该商品所在零售超市的 logo 图标,可以让用户知道商品的售卖来源,下方显示的是商品的名称以及商品的价格,在浏览时该商品的主要信息就可以让用户清楚的知道,避免用户不停地在进入该商品界面和退出该商品界面之间来回切换。



图 6-1

主页面的左上角图标是商品的分类界面,点击该图标后显示的是商品的分类栏目。商品的分类方式是按照商品本身的属性来进行的,在第一级分类中,分为水果、蔬菜、海鲜、肉类、日用品、饮料等;在第二级分类方式中对不同的商品进行了细分,如水果分为特价水果、当季水果、进口水果、本地水果等;之后是最后一个层级,显示的是单指的某种水果或蔬菜、海鲜的名称,例如在进口水果中,分为了奇异果、椰子、芒果、菠萝蜜等,在选中了某种特定的水果后,有关该水果的所有的商品就可以被全部搜索出来。如图 6-2 右图所示,菠萝蜜的商品只有在沃尔玛和家乐福有售,并且只有单一的品种,如果商品的数量不足以填满整个页面,那么之后的商品会由系统随机抽取,显示出系统推荐的其他生鲜食品的商品信息。



6-2a

6-2b

图 6-2

6.2 个人中心

进入个人中心后,首先是用户的登陆界面,如果用户第一次登陆,没有账号和密码,可以点击“注册”,进入信息的注册界面,填写用户名和密码进行个人信息的注册,该应用程序的键盘以及输入法使用的是苹果的 IOS 系统;如果用户忘记密码后,可以进行密码的重置,只需要验证过相关的个人信息后就可以进行密码的重置。如图 6—3 所示



图 6—3

登陆个人界面后,用户可以查看个人浏览的相关信息,在界面的左上角显示了个人的信息,包括个人的预设头像,以及个人的购买信用等级。商家也可以此来判断用户的个人过往购买信誉,有一定的相互监督、相互制约的作用。在右上角的设置图标中,用户可以更改个人的用户密码,改变界面背景的图片等。在界面的主页面上,有个人浏览历史,订单的查询以及售后的相关服务的信息。如图 6—4 所示

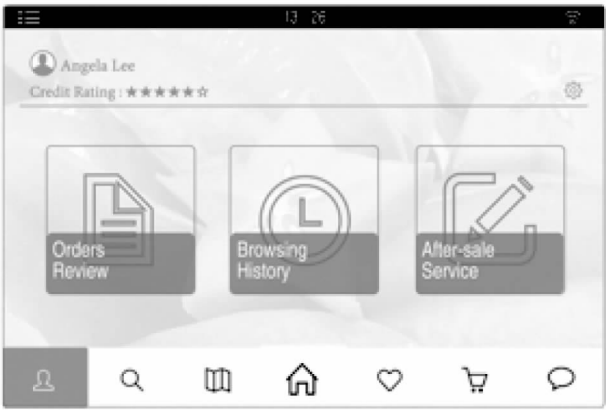


图 6—4

在订单查询中,用户可以查看以往的个人订单,也可以进行分类的浏览,包括已完成的订单,正在配送中的订单,以及需要付款的订单。用户依次可以了解自己在一定时间段内的购买记录。

在浏览历史中,用户可以查询自己在一定时间内的浏览记录,默认的时间段是二十四小时之内的,用户也可以自行进行选择,选择查看一个月内或是半个月内的浏览记录,在浏览记录中如果用户发现需要购买的商品,用户可以进入商品的详细介绍界面,之后选择加入购物车或是进行商品的收藏。

在售后服务中,提供了两种服务,换货和退款,用户在确认订单后,如发现商品有质量问题或是新鲜度等方面的问题,用户可以进行退款或者换货的服务,由于生鲜食品的特殊性,保质期时间较短,因此商家在处理退货或者换货的限制时间不能超过一个规定内的时间,否则时间过长容易引起商品的变质腐坏,如果限定的时间内商家没有做出任何回应,则默认商家同意用户进行换货或者退款的要求,用户可以采取电话上门取货等方式进行原有商品的退换。在商品退换过程中,用户可以在退换货的界面查看退换货的进度以及实时的状况。这样会让商家对于其所购买的商品更为放心,由此对于零售商家来说,也有益于其营业额的增长。此外,在退换货的界面中,商家也可以查看个人以往的退换货订单记录,对于相应订单的时间、商品明细、金额等都可以进行较为细致的查询。

6.3 搜索界面

搜索界面是用户进行单一商品搜索的地方,如果用户有着很明确的购买趋向,用户可以通过搜索该商品名称得到该商品的相应商品信息,在商品进行搜索后,得出有关该商品的所有相关信息,用户可以选择商品信息的排序方式,排序的依据有价格、信誉和商品销售量等划分,通过用户选择的排序方式,用户得到相应的商品信息,就可以对商品进行浏览。选择某一商品后,进入商品的详细介绍界面,包括该商品的大众评价、销售记录、详细图文信息等内容,用户可从这三个方面对该商品进行详细的了解,之后选择加入购物车或是直接购买,抑或是先暂时加入收藏。加入购物车或是直接购买时,在本界面选择需要购买的商品数量以及大约的商品价格。如图 6-5 中右下角的图所示。

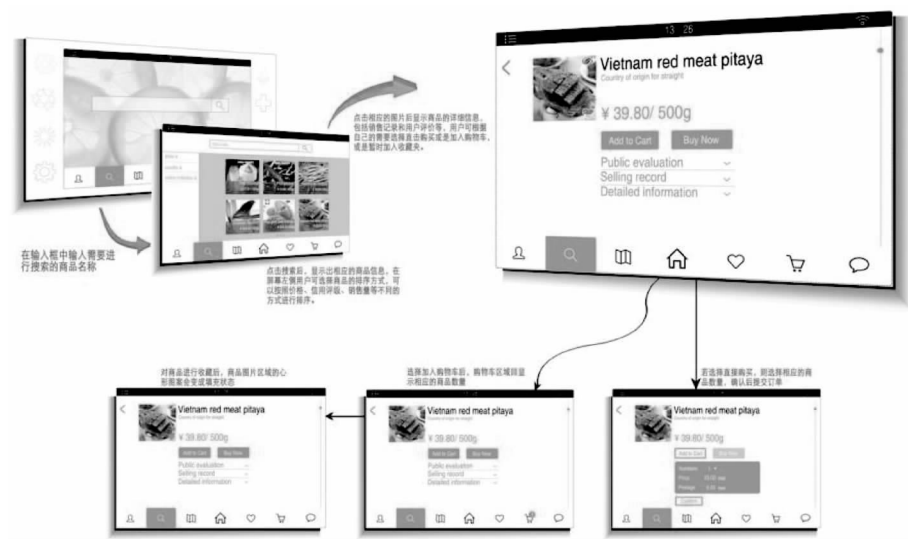


图 6-5

6.4 食谱界面

食谱界面的设置的初始原因是,由于越来越多的年轻人开始独立生活,却又缺少独自烹饪的经验,

因此导致所作菜肴难以入口,久而久之,就习惯于购买外面的食物,但是由于外面食材的新鲜度以及所用调味品质量问题,长时间食用外带食品,对人们的身体健康不利,生活成本也会加大,对于很多上班族来说并不现实。因此,食谱一栏可以让人们在做菜的时候有所参考,做出更多色香味俱全的菜肴。由于大部分家庭的冰箱都放置在厨房内,因此食谱可以在用户烹饪时很方便地加以使用。此外,用户在购买食材时,也可以根据想要烹饪的菜肴的食谱进行相应商品的选择,这样有目的地选择食材,就不容易出现食材浪费或被遗忘等结果。另一方面,食谱的分类是按照原有食材的种类来进行分类的,这样就有目的地避免了与人们饮食习惯上的冲突,例如某些素食主义者就可以在一级分类界面中选择蔬菜一栏,而某些拥有民族习惯的人如果在饮食上有某些方面的忌讳,也可以避免有相应食材的掺入,以免引起用户的生理或心理方面的反感。在食谱的第一层级中,按照食材本身的属性分为蔬菜、肉类和海鲜类,在第二层级中,以肉类为例,按照肉食品的不同种类,分为鸡肉、牛肉、羊肉、猪肉等,在猪肉一栏中,显示的是第三层级的信息,即使用猪肉材料做出的不同菜肴。此次应用程序的界面展示以猪肉中的糖醋排骨为例,按照食材准备、调味品准备、制作方法等顺序对烹饪的方法进行了详细的阐述。在烹饪详述界面的图片下方有两个图标,心形图标表示用户如果对该食谱的菜肴较为喜欢,希望将来易于查看,那么可以进行收藏;另一个图标是视频播放图标,某些用户并不善于阅读食谱,对于文字和图片组成的食谱并不敏感,因此视频上的烹饪演示可以让用户更为明晰地了解该菜肴的制作过程和方法,避免理解上出现误差,导致出错。



图 6—6

6.5 收藏界面

收藏界面分为两个部分,一部分是对商品的收藏,另一部分是对食谱的收藏。收藏的图标用一颗心来表示,当用户在浏览商品或食谱时,如果想要将该商品或是食谱收藏到收藏夹内,那么单击空心的图标,当空心的图标变成实心的图标,就说明该商品或是食谱已经成功加入收藏夹内。用户就可以在收藏栏里进行相应的搜索和查看。

商品的收藏部分中,如果用户看中了某种商品,但是暂时不需要购买,可以将该商品加入到收藏里,这样便于以后再需要该商品时,及时查找,在收藏夹内的商品,用户长按的方式可以点选商品,选择商品后,用户可以选择删除该商品,或是加入购物车,或是直接支付,界面的右上角会显示所选商品的总额以及商品的数量,用户在确认无误后,可以在收藏的界面里直接下单进行购买支付,因此更为便捷。单击商品的图片可以直接查看该商品的详细信息,用户就可以减少界面交互的次数,由此可以使用户使用起来更为方便自然。



图 6—7

在食谱的界面里，显示的是用户曾经进行收藏过的食谱。在食谱少于二十份的时候，用户的收藏食谱界面是系统默认的界面，即类似于商品显示的界面，并不对不同的食谱进行归类；当食谱的数量多于二十份以后，系统自动按照食谱界面第一层级的分类方法对其进行分类；由于第一层级是分为水果、蔬菜和生鲜食品，当其中的任何一项的食谱数量多于二十份时，该层级的食谱会依据食谱内第二层级的分类方法对其进行归类。这样使得食谱收藏的界面看起来更为系统化，更易于查找，因此也就更方便用户进行相关的操作。

6.6 购物车界面

用户在选择商品后，决定购买时，有两种方式，一种是选择商品的数量后将商品加入购物车，这样



图 6—8

就会转到购物车的界面,在购物车的界面内,用户可以选择之前和现在加入购物车的商品,选择完成后,界面的右上角会显示所选商品的所有的数量和总共的价格,用户在确定无误后,就可以进行订单的确认;之后转到订单个人信息填写的界面,用户在这里填写个人的收货信息,例如联系方式,地址确认,姓名,以及配送时间等,此外,用户可以在备注一栏里添加额外的录入信息,完成后,用户就可以下单,选择支付方式,支付成功后,就可以等待商品送货上门了。另一种是选择商品的数量后直接购买,这样就省去了购物车的界面,直接进入到了订单信息填写的界面。

在购物车的界面,商品的最大存储量为三十个种类的商品,如果超过这个数量,那么商品就没办法再进行录入。此外,由于生鲜食品本身的特殊性,很多的特价信息以及商品的保持时间是更换较为频繁的,例如在四月初有的水果可能到了五月底就不再进行售卖,甚至在之后的半年多的时间里都不会出现该商品的售卖信息,针对这一类的商品,系统会在商品取消售卖后自动将该商品进行删除,并且会在用户的个人信息一栏里给予信息的提醒。此外,用户也可以对商品的逾期时间进行设置,例如将芒果这种季节性较强的水果设置提醒时间为十二小时,如果商家预期在十二小时后该商品就会售卖一空,并且不会再有进货的来源,那么商家就可以对设置了提醒的用户进行消息的提醒。这样有两方面的优点,一方面用户不会因为遗忘或是犹豫不决等原因错过心仪的商品,另一方面,如果该商品已经是最后一批,商家进行提醒后,销售量一般会有预计性的攀升,商品可以尽快的出库,这样可以保持商家库存的流动性,而不会使得最后一个批次的商品面临占用库存却又可能因腐坏而扔掉的结果。

6.7 推荐圈界面

推荐圈界面是一个流动性的信息墙,用户可以在这里相互交流对于食材营养的看法,食谱的经验,商品的推荐等一系列的信息,这里的信息都是实时更换的,以时间为依据,当新出的推荐信息出现后,最后的信息就会自动删除,不再保存。如果用户在浏览过程中发现认为有用的信息时,可以点击界面右上角的心形图标,进行商品的收藏。此外用户也可通过大众常用的交流软软件,将相关的信息文章转发给朋友或是进行分享。

推荐圈的界面有两方面的意义,一方面是通过相互交流的方式,可以让人们对于生鲜食品的膳食营养变得更为有趣,并引起相应的重视,由此更为有益于改善人们的身体健康状况;另一方面,如今主流的交流平台关于膳食营养的相关信息太过贫乏,而人们很少会主动查询这方面的相关信息,因此这



图 6—9

样一个交流空间可以使人们在无意识的情况下汲取更多的营养膳食方面的信息,此外,在膳食调理方面,有很多是个人家庭的传统方法,或是基于个人经验进行的总结,这一类的信息一般不会在网络空间内查询得到,因此这样一个交流的空间给了人们一个有效的平台让人们分享自己的经验,获知他人的经验,更好地实现健康饮食,健康生活的目的。

7 结论

7.1 研究总结

本文是基于智能冰箱为载体,进行了一种新的销售模式的研究,由此来应对在电子商务冲击下,传统零售行业节节败退的尴尬局面。另一方面也实现了生鲜食品等具有保质期较短,不宜碰撞的特点的商品的不出门购物,使得人们的生活变得更为便利,也通过食谱、营养搭配等信息提供以作参考的方式来使得人们的生活变得更为健康,改变现今社会中,人们在面临巨大社会压力的情况下,身体状态长期处在亚健康的生活状况。

本文的创新点主要有三个方面:

1. 智能冰箱的定义性拓展。“智能冰箱”的概念已经越来越频繁地进入到大众的视野,随着技术的不断发展,人们对于电子产品的智能化有着越来越高的要求,智能化的实现度也越来越高。智能冰箱现如今已经实现了变频制冷,搭载安卓系统,扫描商品管理商品保质期等一系列的功能。但是现如今冰箱的智能化还是停留在其最基本、最初始的功能上——制冷,保鲜。这使得冰箱的智能化发展本身也受到了一定的局限性。未来的时代是信息化的时代,所有的电子产品或是电器都会向信息化的方向进行发展,而不是只满足于其初始的物质功能上的需求。因此,本次智能冰箱的交互设计发展研究将冰箱的智能化扩展到了另外一个全新的领域,使得冰箱的智能化更为深入,也更为深入地渗透到人们的日常生活中,将其功能扩展到了更外一个层面,承担了定义上更为广泛的功能责任。

2. 对当今社会压力下人们身体健康状态的重视改变。科技的快速发展使得人们的生活越来越便捷,但是有另一方面的影响,就是人们的生活节奏明显的加快,生活节奏的变快使得人们对于自己能完成的价值期望越来越高,在某种程度上甚至超越了人类的身体机能所能承受的极限点,由此,在最新的报道中,年轻人由于饮食不规律、工作时间过长或是身体体质较差等一系列的原因造成过劳死的后果的新闻层出不穷。人们越来越以追求极限的方式来追求效率,却忽略了对本身身体健康的关注,这是经济与科技过于飞速的发展所带来的负面效果。因此在智能冰箱的交互设计中,这种新的零售模式一方面可以让人们有选择性地对生鲜食品的采买,既可以出门购买来放松心情,也可以在室内购买来节省时间做自己更愿意做,有利于放松精神心情的事情,而不是为了生理上的需要要疲于奔命地必须花费相当的时间来满足自己的饮食需求;另一方面,在应用程序中提供的一系列的关于食材具体营养成分,饮食营养搭配,食谱推荐等参考信息,也可以让人们对于自己的身体健康状态更为重视,从而达到改善人们现有的普遍亚健康的身体状态的状况。

3. 为传统零售商寻求新的发展途径。自进入二十一世纪以来,电子商务的发展飞速增长,在经过本世纪初的互联网冬天后,电子商务的发展更是可以用“势如破竹”来形容,在抢占传统市场的营业额方面,电子商务基本没有遇到太大的阻碍就抢占了大量的市场份额。究其原因是因为中国市场特有的几个特点:一是中国的制造业发达,并且在遭遇全球金融危机后,由于人民币的升值,许多小制造商将目光从国外市场转向国内市场;二是中国的小微型企业太多,但是面临的市场环境却是极端恶化,创业的条件和发展的条件都不甚成熟,市场环境也不甚乐观,而电子商务的创业平台却不需要投入太大的

成本就可以进行运营,因此在短时间内电子商务就积累了相当大的一部分用户群体;三是中国由于房地产行业的畸形发展,导致各方面的土地成本激增,由此相关的商务用地租金也猛增,这样导致商品的在实体经济的大环境中成本过高,与之相对的电子平台的商品就因此有了巨大的价格优势。基于以上三个原因,传统的零售模式想要对抗电子商务的冲击,情况不甚乐观,并且截止到现在为止,很多传统的大型零售商,例如家乐福、沃尔玛等都已经开始了网络商店的经营,或是搭载像一号店这样的电子商务平台来进行合作运营,但是效果都不是非常的理想。而这种新的零售模式充分发挥了传统零售商的既有优势,所需要付出的转型成本有微乎其微,因此就使得其有了与电子商务平台竞争的条件,可以有效地改善现如今被动的尴尬局面。

7.2 问题与展望

本文已经取得了一定的成果,例如对于此种商业模式的市场调查,所调查的用户人群百分之九十以上的人对于这种商业模式抱有期待和乐观的态度,认为它的实现对于人们生活的便利度和健康度都有着明显较大的改善。所面临的问题就在于这种商业模式的转型需要尝试的零售行业承担一定的商业风险,并且因为这种商业模式的成功与否在于人们现有的购买生鲜食品的习惯是否能否有所改变,而整个社会的人们的生活习惯的改变是一件风险性较高的投资,因此此种零售模式的它的实现度以及效果有着强大的不确定性。

参 考 文 献

- [1] [美]科尔科(Kolko Jon),方舟译.交互设计沉思录[M].北京:机械工业出版社,2012年10月
- [2] 陈星.基于网络平台的智能冰箱应用探讨[J].北京:科技创新导报,2013年3月
- [3] 董茂,王志恒,张艳杰,旬一,杨庆华.基于WiFi网络的成套商用冰箱智能控制器研究[J].杭州:机电工程,2014年1月
- [4] 黄贤强.交互设计在工业设计中的应用研究[J].济南:齐鲁工业大学,2014年6月
- [5] 吴瑜.人机交互设计界面问题研究[J].武汉:武汉理工大学,2004年5月
- [6] 李世国,华梅立,贾锐.产品设计的新模式——交互设计[J].武汉:包装工程,2007年4月
- [7] 李四达.交互设计概论[M].北京:清华大学出版社,2009年9月
- [8] [英]Giles Colborne.简约至上:交互式设计四策略[M].北京:人民邮电出版社,2011年1月
- [9] [美]Susan Weinschenk,徐佳,马迪,余盈亿译.设计师要懂心理学[M].北京:人民邮电出版社,2013年5月
- [10] 腾讯公司用户研究与体验设计部.在你身边,为你设计 腾讯的用户体验设计之道[M],北京:电子工业出版社,2013年1月
- [11] 王月茹.基于产品形象的厨电产品交互界面研究[J].南京:南京理工大学,2011年12月
- [12] [美]诺曼,付秋芳,程进三译.情感化设计[M].北京:电子工业出版社,2005年5月
- [13] 日本奥博斯科编辑部,暴风明译.配色设计原理[M].北京:中国青年出版社,2009年12月
- [14] 杭间.设计道——中国当代设计艺术的基本问题[M].重庆:重庆大学出版社,2009年1月
- [15] [日]佐藤可士和.佐藤可士和的创意思考术[M].北京:北京科学技术出版社,2011年8月
- [16] [日]原研哉,纪江红译.[M].南宁:广西师范大学出版社,2012年3月
- [17] [美]加瑞特,范晓燕译.用户体验要素:以用户为中心的产品设计[M].北京:机械工业出版社,2011年7月
- [18] 林影丽.冰箱界面人性化设计研究[J].天津:天津大学,2005年12月
- [19] 搜狐新闻客户端UED团队.设计之下——搜狐新闻客户端的用户体验设计[M].北京:电子工业出版社,2014年1月

- [20] 黄琦,毕志卫. 交互设计[M]. 杭州:浙江大学出版社,2012年7月
- [21] [美]库伯,瑞宁,克洛林著,刘松涛译. About Face3 交互设计精髓[M]. 北京:电子工业出版社,2012年3月
- [22] 姜坤鹏. 移动电子商务交互设计研究[J]. 北京:首都师范大学,2013年5月
- [23] 郑杨硕. 信息交互设计方式的历史演进研究[J]. 武汉:武汉理工大学,2013年4月
- [24] 王玉珊. 由情感记忆引导的产品交互设计研究[J]. 无锡:江南大学,2012年6月
- [25] 牛纯. 新媒体展示设计中的交互设计研究[J]. 北京:北京印刷学院,2014年12月
- [26] 孙传祥. 基于自然交互方式的移动终端产品交互设计研究[J]. 济南:山东大学,2012年4月
- [27] 王冬. 现代生活产品中产品的人机交互设计[J]. 无锡:江南大学,2012年4月
- [28] 马颖云. 购物比价网站的交互设计研究[J]. 上海:华东理工大学,2013年12月
- [29] 舒昕. 购物网站受众本位交互设计研究[J]. 湘潭:湘潭大学,2012年6月
- [30] 刘超. 基于隐喻理解的移动终端界面交互设计[J]. 北京:北京邮电大学,2010年3月
- [31] 陈媛媛. 基于活动的情境感知模型与情境感知交互设计[J]. 大连:大连海事大学,2012年12月
- [32] 郝冀宣. 现代家电产品交互界面设计研究[J]. 南京:南京航空航天大学,2013年3月
- [33][日]黑川雅之,吴俊伸译. 素养与身体[M]. 石家庄:河北美术出版社,2013年9月
- [34] 赵大羽,关东升. 交互设计的艺术——iOS 7 拟物化到扁平化革命[M]. 北京:清华大学,2014年2月

外文参考文献

- [1] Bill Moggridge. Designing Interactions [M]. Cambridge,Massachusetts: MIT Press (MA),2006年11月
- [2] Luke Wroblewski. Web Form Design: Filling in the Blanks[M]. Brooklyn, New York: Rosenfeld Media,2008年
- [3] Jesse James Garraett. The Elements of User Experience[M]. San Francisco, California: Peachpit Press,2002年10月

The Research on the Future Refrigerator and its Interaction design

Qi Wenjing

(School of Electronic and Information Engineering, North China Univ. of Tech. , Beijing, 100144)

Abstract Although the refrigerator develops more and more rapidly, the character of the refrigerator is still confined as a tool which can be used to keep fruits and vegetables fresh. Which means, it doesn't shows the intelligent aspect in the information age. As this paper argues, the relationship between people and refrigerator is the critical point on touching the real intelligent in the information age. On the other hand, although the most commodities can be sold in a way of electronic business, the fresh food can't which because it can be collided easily and can not be keeping fresh for a long time. , which means it can't be transported in a long distance. But in the stresses and strains of today's fast-paced society, a lot of people have less time on buying the commodities and fresh food by themselves, which indicates the great inconvenient for them. Therefore, realizing the shopping on the fresh food in the electronic business can bring the great convenience for people.

This paper analyst the numbers and current situation of the traditional retailers and the electronic commerce in recent years. Then analyzing the existing intelligence refrigerator which had been produced in the market, analyzing the possibility on the connection between wifi and refrigerator. Based

on these analyze, at last, designing the interactive interface of the app in the people orientation, function and color matching, which determine seven parts' interactive design and interactive mode, such as the personal center, search, cookbook, home page, favorite, shopping cart, communication part. The design of the interactive surface and the systemized of the whole shopping process could be the most important work in this paper.

Key words intelligent refrigerator; sales model, selling; e-business; interactive design

北京地区高密度居住区户外活动空间多功能利用研究

研究生 王 昭

导师 卜德清

(北方工业大学建筑与艺术学院 北京 100144)

摘 要 近些年,北京地区呈现出人口不断密集同时用地趋于有限的状况,因此如何在高密度开发模式的居住区中合理解决场地与功能之间的需求问题,或怎样更好地达到两者之间的平衡是本文研究的关键。本文选取北京典型的高密度居住区国典华园、华清嘉园、富润家园为主要研究对象,运用环境行为学的理论和方法进行现场调查,对使用人群、活动类型进行划分,采用统计法对不同人群的行为过程进行详细记录,在此基础上对场地的功能使用情况进行归纳,深入分析围合性、遮阴、休息设施、尺度、区位等因子对高密度居住区户外活动空间的功能影响,探讨这些可能的影响因素与场地活动频繁程度之间的关系,总结出影响因素的差异对场地利用率的影响,从而得出高密度居住区多功能户外活动空间的设计方法,为居住区的多功能活动场地设计提供参考。

第一章介绍了在北京人口密集和用地有限的背景下,提高高密度居住区户外活动空间利用率的迫切性。目前对居住区的场地及空间的研究有一部分提到场地需要多功能化的建议,还没有具体的一篇文章阐述如何对场地或空间的多功能进行研究,此处为本文的创新点。

第二章对“高密度居住区”、“户外活动空间”和“户外活动空间利用率”的概念进行界定,对户外活动空间构成要素进行阐述,并对应用到的外部空间设计理论进行介绍。

第三章阐述了居住区户外活动空间具有限制性、公共性和领域性的特点以及户外活动空间多功能性的三类表现方式:按使用人群、按时段、按活动类型,并对居住区出现的活动类型进行分类,方便后面章节的分析研究。第四章是本文的重点章节,主要是对北京典型高密度居住区户外活动空间进行的调研与分析。通过对高密度居住区户外活动空间的整体设计、预期功能、人员活动情况、空间功能变化与特征等不同方面来统计分析,得到相应的调研数据,分析可能对功能产生影响的具体因素,总结出利用率相对高的活动场地所具备的特征。

第五章在可能对空间功能产生影响的相关因素的调研基础上,深入探讨居住区的围合性、遮阴、休息设施、尺度、区位等因子在创造多功能户外活动空间过程中的影响,同时总结出不同活动类型对场地的不同要求,提出功能适当融合、划分空间层次、增加细节处理、多功能的控制等设计要点。

第六章为本文的结论,例如两面围合与三面围合理论上应作为多功能场地主要的围合方式;主要道路与场地的关系建议为“空间一体化”; $140\text{m}^2 \sim 200\text{m}^2$ 左右的中型空间是本文推荐的多功能场地规模等等。

关键词 高密度居住区;户外活动空间;多功能;研究

作者简介 王昭,女,1988年出生,硕士研究生,主要研究方向为环境景观设计。

E-mail: familyzhao_0906@126.com

1 绪论

1.1 研究背景

高密度开发使居住区户外活动空间质量变差,活动场地趋向紧张,居民的活动空间变得局促,由于

人口密集和有限的用地使人们居住的户外环境质量无法得到提高,部分活动需求无法满足。因此,如何能在高密度居住区的建设运行过程中,保持和改善居住区景观环境质量,使有限的户外活动时间发挥它的最大作用,为居民提供舒适的户外活动时间,是目前一个十分重要的问题。

1.2 研究目的及意义

通过对北京地区高密度居住区户外活动空间的调研,总结现有活动空间的功能和特征,提出在空间限定、道路与场地关系、遮阴、休息设施、尺度、区位等因素上影响活动空间的多功能性,试图找到多功能活动空间的内在规律与设计方法,进而指导未来的多功能活动空间设计,提高活动空间的利用率,优化居住区户外活动时间质量。

1.3 国内外研究现状

国外对空间的研究很多,如《美国大城市的生与死》,扬·盖尔的《交往与空间》,克莱尔·库珀·马库斯的《人性场所》等。

国内对多功能的研究在土地与建筑空间较为多见,对户外活动时间空间多功能利用研究较少,目前对居住区的场地及空间的研究不在少数,也有相当一部分文章提到场地需要多功能化的建议,但还没有具体的一篇文章阐述如何对场地或空间的多功能进行研究,宏观的建议到具体的实施方法还是有很大的研究空间的。

1.4 研究内容及方法

通过文献研究、实地调研、定性与定量分析、微观与宏观分析相结合的方法,同时运用环境行为学、环境心理学、统计学等多学科理论对行为与户外活动时间空间进行综合分析。

2 正文

2.1 北京高密度居住区户外活动时间空间多功能利用的调研与评价

以华清嘉园为例:

(1)调研概要:华清嘉园容积率 2.5,建筑密度 27%,层高 14—20,绿化率 30%,楼房属于自由式布局,户外活动时间空间较为分散,按照空间顺序将华清嘉园中的户外活动时间空间分为 1~3 号场地。



图 4—10 华清嘉园区位图

(图片来源:作者自绘)

(2) 活动空间整体设计分析及预期功能

1号场地位于小区内主要交通一侧,与周边有一定的高差变化,西侧与一儿童游戏场相连,空间边界以花池种植为主,内部种植形成小树阵,为夏季遮阴提供了条件,边缘布置少量座椅,四面各有道路相通,空间尺度适宜居民进行成群结组地进行活动。1号场地的预期功能为休息与交流、小型成组的健身活动、家长照看小孩。

2号场地与1号场地相连,位于1号场地东侧,空间呈狭长形,用于活动的空间有限,主要以中央水池为观景景观,空间边界为景观灯柱,内部无种植,布置少量座椅,东西侧交通。2号场地的预期功能为休息与交流、小型分散的健身活动。

3号场地呈圆形,空间边界为长廊花架,内向型空间特征明显,内部四个五边形种植池配合木质座椅,此外空间边缘仍布置有其他休息座椅,四面交通,空间尺度适宜三五成群的小规模活动。3号场地的预期功能为休息与交流、小型的健身活动、家长照看小孩。



1号场地现状



2号场地现状

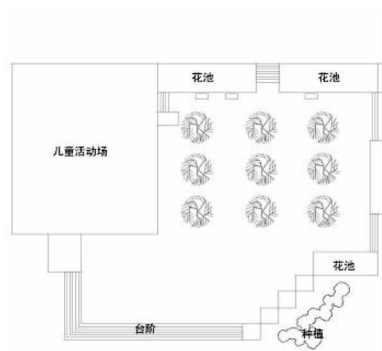


3号场地现状

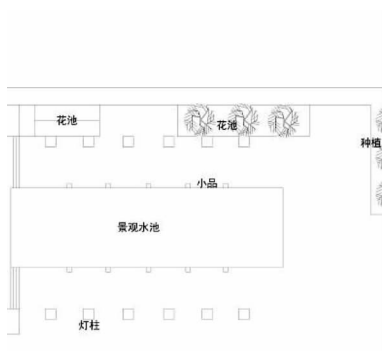
图 4-11 华清嘉园场地现状

(图片来源:作者自摄)

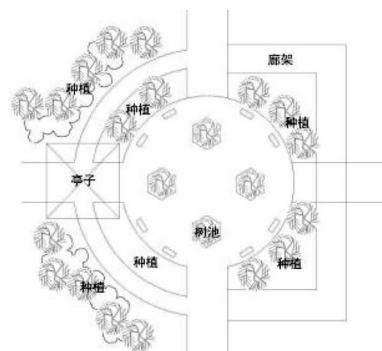
(3) 活动空间人员参与分布情况



1号场地



2号场地



3号场地

图 4-12 活动场地平面图

(图片来源:作者自绘)

健身和锻炼在不同的调研时间呈现的规律基本稳定,在早上7点到9点左右成群的中老年人围绕1号场地做操,占据大部分健身和锻炼的参与人员。照看孩子的活动基本保持两个高峰,一个高峰为

11 点左右,另一个高峰为 16 点左右。休息和交流与其他活动类型的人员数量随时间的变化曲线较缓。



2014. 1. 13 7:32



2014. 11. 3 09:55



2015. 1. 18 11:12



2014. 12. 19 14:36



2014. 11. 3 16:03

1 号场地人员活动情况



2014. 12. 19 9:28



2013. 12. 16 10:31



2014. 11. 3 14:50



2014. 11. 3 15:32



2015. 4. 9 16:02

2 号场地人员活动情况



图 4—13 华清嘉园户外空间人员活动情况

(图片来源:作者自摄)

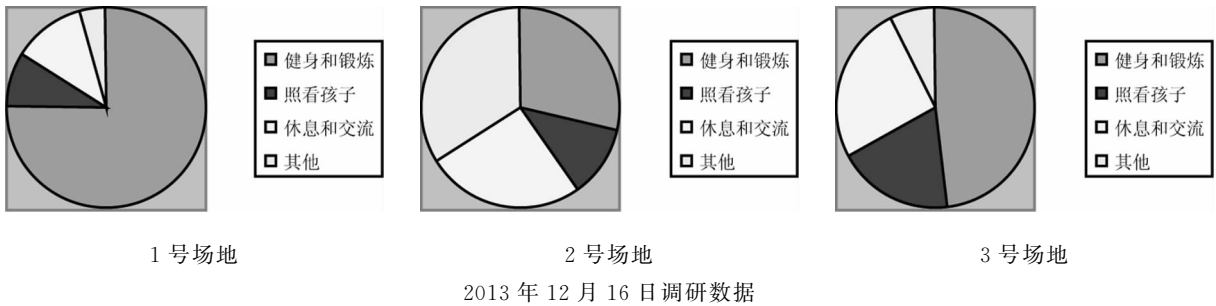
1号场地人员分布主要聚集在休息设施附近,其次是与儿童活动场地相接的边缘或坡道附近,穿行人流主要为与2号场地发生关系的双向人流,目的是汇向主要交通,外部人流通常从场地南侧汇入主要道路,对1号场地影响较小,因此健身和锻炼等活动仍可以发生。

2号场地人员停留较少,较为分散,穿行人流占使用人群的大多数,由于景观空间的独特性使得穿过式交通严重影响场地的使用。

3号场地人员分布主要聚集在北侧入口的座椅处,这里与主要入口相连,方便停留,穿行人流较少,大部分人流从西侧道路绕行,少数人从树池边穿行,东侧花架基本无人活动。

(4)活动空间功能变化

调研中各场地各类活动所占的比例如图所示。



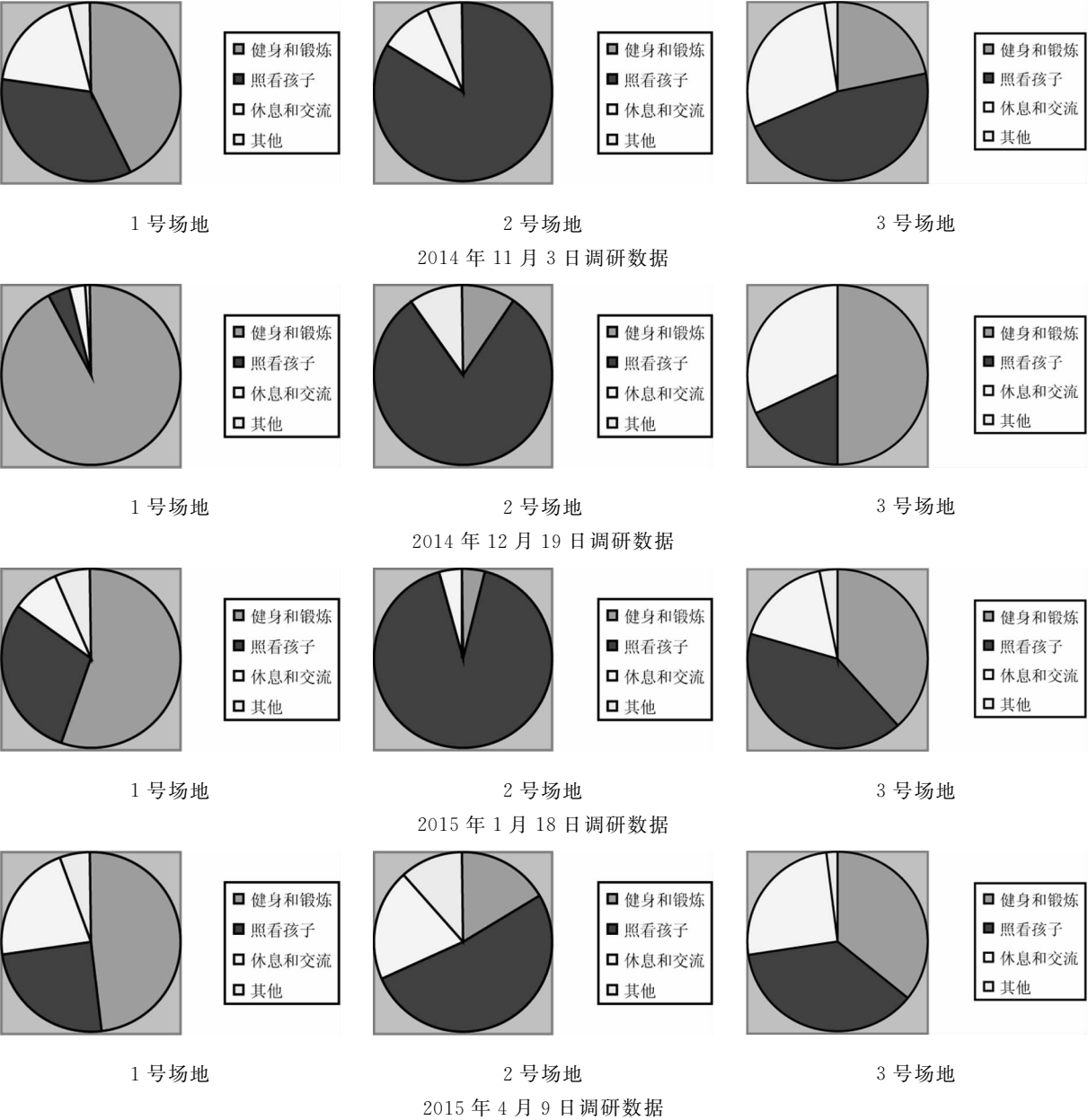


图 4—14 各场地不同活动类型所占比例

(图片来源:作者自绘)

从上表可以看出:

- 1 号场地其开敞的空间特点决定发生在这里的活动类型主要为健身和锻炼(平均调研数据为 66.2%),且集中在早上时段。
- 2 号场地是以景观为主的观赏空间,除了第一次调研发生的主要活动为经过、打电话等其他活动类型(调研数据为 34%)外,发生的活动类型主要为照看孩子(在景观水池中踢球),总的平均调研数据为 66.6%。
- 3 号场地为围合型空间,发生的主要活动类型为健身和锻炼(调研数据为 39.3%),其次是休息和交流(调研数据为 31.2%)。

表 4—13 华清嘉园活动场地预期功能与实际功能对比

场地编号	预期功能	实际功能
1 号	休息与交流、小型成组的健身活动、家长照看小孩	健身和锻炼、休息和交流、照看孩子、其他(遛狗、打电话)
2 号	休息与交流、小型分散的健身活动	健身和锻炼、休息和交流、照看孩子、其他(路过、打电话)
3 号	休息与交流、小型的健身活动、家长照看小孩	健身和锻炼、休息和交流、照看孩子、其他(看报)

与预期功能出现差异的主要为 2 号场地,初期认为景观水池的高差与无防护措施会不利于照看孩子的活动类型发生,但调研发现,有一些家长会分散在场地边缘与水池距离较远处、利用相对空间较大处进行此项活动,说明只要安全尺度足够大,不良因素的影响会被削弱。同时调研后期发现一些学龄儿童经常在不注水的水池中踢足球,水池的边界限制反而为他们提供了良好的游玩条件。

(5)活动空间特征

表 4—14 华清嘉园活动场地空间限定与边界分析

(图片来源:作者自摄)

空间编号	限定因素	限定面数	限定强度	照片
1 号	高差(台阶、残疾人坡道)	4	强	
2 号	灯柱、植物	2	较弱	
3 号	廊架、矮墙	4	强	

1 号场地通过高差进行限定空间范围,其间无其他限定方式,因此在这里发生的活动多为对尺度

要求较高的成组群形式的锻炼。2号场地除了外围的种植外,在空间内部以灯柱进行二次限定,但由于小品的体量、场地的景观布置,限定效果不明显,活动较分散。3号场地通过廊架、矮墙等小品加强圆形的边界限定,并且场地中具备舒适的荫凉,休息设施的布置使其利用率更高,这种限定使发生在这里的活动呈现多样性。

表 4—15 华清嘉园活动场地主要道路与场地关系分析

(图片来源:作者自绘)

空间编号	主要道路与场地关系	道路对场地影响	连接场地道路宽度	图示
1 号	相接	较小	3m	
2 号	穿过	大	3m	
3 号	相接	较小	1.5m	

1号场地位于小区内主要交通一侧,与周边有一定的高差变化,西侧与一儿童游戏场相连,四面各有道路相通,属于相对开放的活动空间。活动空间与交通空间关系为相接。2号场地与1号场地相连,位于1号场地东侧,空间呈狭长形,以中央水池为观景景观,东西侧交通,虽然2号场地同样位于主要道路一侧,但其空间构造使得内部人流占据主要交通地位。交通空间穿过活动空间。3号场地呈圆形,被廊架与矮墙围合,内向型空间特征明显,四面交通,但以北面与西面交通为主。活动空间与交通空间关系为相接。

表 4-16 华清嘉园活动场地遮阴分析

空间编号	绿化与场地关系	绿化形式	遮阴效果
1 号	位于场地内、周边	乔木、灌木	好
2 号	位于场地外	乔木	较差
3 号	位于场地内、周边	乔木、灌木	好

1 号场地由于其开敞的空间和丰富的植物遮盖条件非常适宜人们晨练的需求,故夏季上午时段在这里进行身体锻炼的人们居多。3 号场地的遮阴条件也属于比较好的,除了场地周边的基础绿化遮阴外,在场地内部的铺装上也有绿化的设计,规则的六边形树池布置,自然形成了理想的林荫场所。2 号场地则缺少场地中的遮阴,因此阳光强烈的夏季在此活动的人们很少,中间形成空置的场所,只有少量利用周边绿化进行遮阴的休息的人们。

表 4-17 华清嘉园活动场地休息设施分析

空间编号	形式	数量	单位面积休息设施数
1 号	定制座椅、花池	3	0.006
2 号	定制座椅	2	0.004
3 号	定制座椅、廊架、矮墙	8	0.071

1、3 号场地都具有基础的座椅,且具备功能性的休息设施,如高于地面 400mm 左右的花池、矮墙等,为休息、照看孩子等活动类型提供条件,空间的利用率相对提高。2 号场地只具有简单的 2 把定制座椅,因此这里的活动受限,像休息、聊天等对休息设施要求较高的活动类型极少。

表 4-18 华清嘉园活动场地空间尺度分析

空间编号	具体尺寸	认识感受
1 号	25mX20m	开敞
2 号	35mX15m	深远
3 号	12m 为直径的圆	围合

1 号场地空间尺度大约为 25m×20m,由于平整的地面和相对开敞的空间,有利于体育活动的开展,进行体育锻炼的人员较为密集。2 号场地空间尺度大约为 35m×15m,空间过于狭长,缺少足够的活动场地,大都被水池、小品等占据。3 号场地空间尺度大约为 110m²,是适合人们交流和活动的舒适空间,人员相对较多。

表 4-19 华清嘉园活动场地区位分析

空间编号	与建筑阴影关系	周边影响因素	空间活性
1 号	场地北侧有高层建筑,位置较远	与儿童活动场、中央水景相邻	大
2 号	场地南北侧均有高层,且距离较近	通往其他空间的便捷通道	较小
3 号	四个斜方位角度均有高层建筑	与一个入口较近,可通往居住区西侧	较大

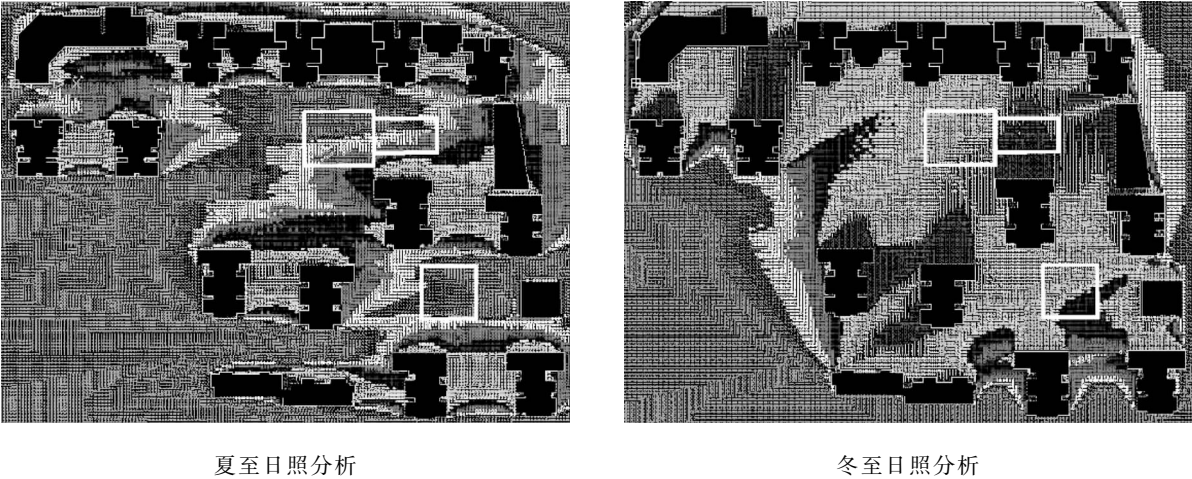


图 4—15 华清嘉园活动场地日照分析

(图片来源:作者自绘)

华清嘉园夏季建筑阴影对活动场地的影响不大。冬至 2 号场地日照时长出现了 0 的情况,同时大部分活动场地处于 0.5~1 个小时的日照时长范围内。说明冬季建筑阴影对 2 号场地产生了较大的影响,出现了长期阴影面,户外采光严重不足,从而影响活动的发生。

(6)小结

表 4—20 华清嘉园场地利用率数据
(2013 年 12 月 16 日)

空间编号	场地面积(m ²)	参与人数	利用率
1 号	500	68	0.136
2 号	525	27	0.051
3 号	113	35	0.310

表 4—21 华清嘉园场地利用率数据
(2014 年 11 月 3 日)

空间编号	场地面积(m ²)	参与人数	利用率
1 号	500	241	0.482
2 号	525	92	0.175
3 号	113	79	0.699

表 4—22 华清嘉园场地利用率数据
(2014 年 12 月 19 日)

空间编号	场地面积(m ²)	参与人数	利用率
1 号	500	90	0.180
2 号	525	10	0.019
3 号	113	18	0.159

表 4—23 华清嘉园场地利用率数据
(2015 年 1 月 18 日)

空间编号	场地面积(m ²)	参与人数	利用率
1 号	500	78	0.156
2 号	525	23	0.044
3 号	113	34	0.301

表 4—24 华清嘉园场地利用率数据
(2015 年 4 月 9 日)

空间编号	场地面积(m ²)	参与人数	利用率
1 号	500	200	0.400
2 号	525	78	0.149
3 号	113	98	0.867

从上表可以看出,场地的平均利用率最高的为 3 号场地,其次是 1 号场地,最低的为 2 号场地。1、3 号场地以健身和锻炼和照看小孩的功能为主,多功能性强于以观赏的功能为主的 2 号场地。

2.2 北京高密度居住区多功能户外活动空间的设计探讨

(1)相关因素分析

空间限定与围合:

通过调研,出现的限定类型有小品和植物限定、铺装限定、高差限定。



图 5—1 限定类型

(图片来源:作者自绘)

调研发现,户外活动空间大都具有围合性,但限定程度有所不同。

(1)四面限定的空间。这种空间封闭性非常强,具有强烈的向心性和内聚性,特别是当这种空间的规模较小时,这种向心性会更强。有时私密性过强,不利于开放性活动的发生,如照看孩子。

(2)三面限定的空间。这种空间的围合感较强,同时也具有一定的方向性和向心性。是较为舒适的空间类型,不仅可以吸引人群参与活动中而且也适宜对环境要求较高的私密性活动如休息和聊天,

不愿受过多他人的干扰。

(3)二面限定的空间。空间限定较弱,空间具有一定流动性。此种空间也适宜多种功能活动的开展,具备了流动的特征,是活性相对较大的空间,人们会聚集在限定面附近,而非限定面则更为灵活地被需要空间较大的活动占用。

(4)一面限定的空间。封闭性很差,但开放性很好。此种空间适宜成组和外向的活动,但多功能性相对较弱,不利于私密性活动的发生。

以上四种情况当限定高度低于 450mm 左右的情况时,限定强度会明显减弱。

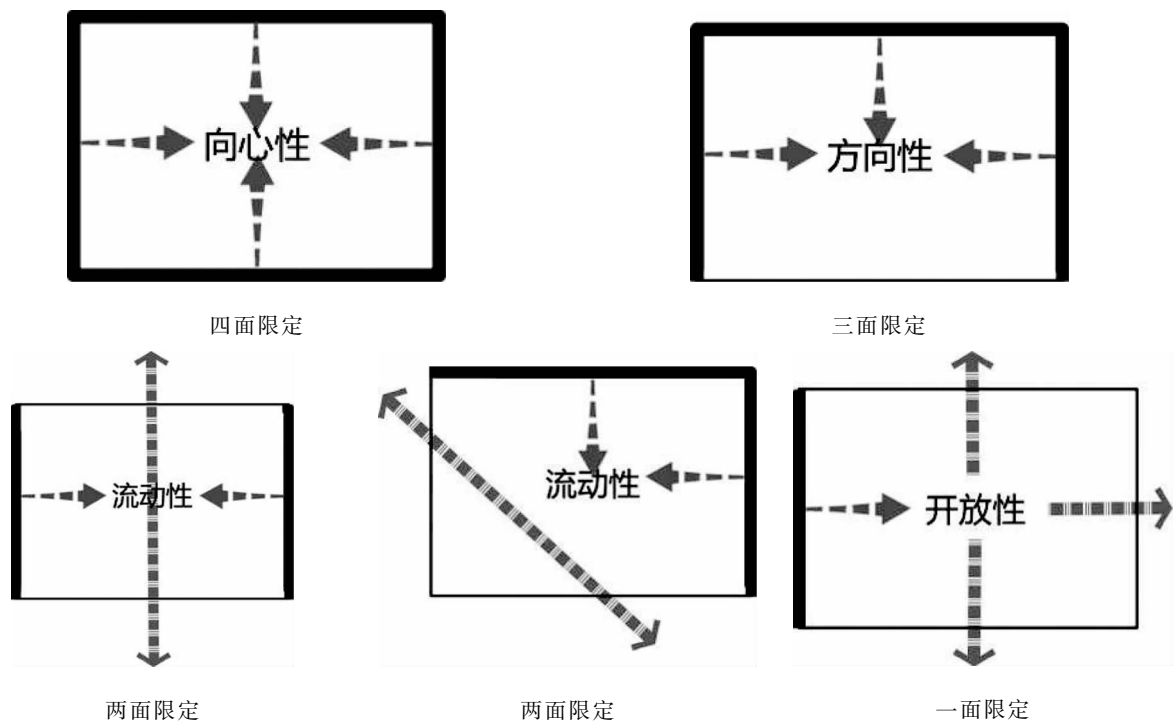


图 5-2 限定面数对空间的影响

(粗线代表限定面,图片来源:作者自绘)

通过统计,所有调研的场地中空间限定类型与面数的数据如下:

表 5-1 调研场地空间限定类型与面数的场地数量统计数据

居住区	空间限定类型			空间限定面数			
	小品、植物	高差	铺装	四面	三面	二面	一面
国典华园	2	0	2	0	2	3	0
华清嘉园	2	1	0	2	0	1	0
富润家园	2	0	2	2	0	1	1
华盛家园	3	1	1	1	2	1	1
锦秋家园	1	0	3	0	0	1	3
逸成东苑	4	2	2	3	2	1	2
总计	14	4	10	8	6	8	7

由上表可以得出,目前高密度居住区内小品和植物的限定空间方式数量最多,其次是单纯的铺装,说明铺装也是限定空间的常用方式,这种限定方法限定的效果不明显,容易形成单一的活动空间,从而影响使用功能,小品和植物、高差类限定方式比例还需提升。虽然植物类限定所占比例较高,但都是以场地周边的植物限定为主,缺少场地内的空间限定与分隔,因此限定效果仍偏向单一。而限定面数在调研中发现占比较平均。

下表通过对其他变量相似,空间限定与围合因素变化的空间进行整理作出比较如下:

表 5-2 调研场地空间限定类型与面数的比较

场地编号	主导活动类型	平均总活动人数	空间限定与围合	
			限定因素	限定面数
国 2	照看孩子	36	铺装、植物	2
华 3	健身和锻炼	53	小品	4
富 1	照看孩子	49	铺装、植物	4
富 3	照看孩子	19	铺装、植物	1

虽然富润家园 3 号场地是四面围合的限定空间,但其限定因素为铺装与植物,所以限定强度有所降低。总的来说照看孩子的活动更倾向于在相对开放的活动空间展开,同时人们更愿意在有明确限定的空间中活动。华清嘉园 3 号场地的四面限定高度低于 450mm,因此仍为相对开放的空间。

主要道路与场地关系:

通过调研总结出主要道路与场地的关系主要有以下三种情况(如图 5-3):

(1)主要交通与活动场地相接—相接模式

在这种模式下一方面道路经过的人与场地的活动之间有良好的视线交流,场地具有良好的公共性和开放性,另一方面道路与场地各自有独立的空间,场地具有一定的领域性,交通不会干扰场地内活动。这是最为理想的模式。当主要道路与场地关系为相离且主要道路与场地距离较近时与相接模式的效果相似。相隔较远但无视线遮挡,利用率仍高于相离较远且有遮挡的情况。

(2)主要交通穿过活动场地—穿过模式

在此种模式下,居住区的居民主要交通流线都会经过活动空间,一方面场地与道路之间有空间共享,场地空间具有最好的公共性和开放性,有利于有更多的人来到场地。但是场地内领域性较差,同时私密性受到侵犯,穿过式的交通会对场地的使用造成一定程度的干扰,活动也会受限。利用率中等。这种模式只有在活动空间尺度很大,主要交通量小的情况下适宜采用,而活动空间尺度小的情况下很容易变成单一的道路空间。

(3)主要交通与活动场地相离—相离模式

在这种模式下,道路与场地距离较远,而且道路与场地之间有绿化遮挡的情况下,场地的领域性很强但是公共性和开放性很差,居民不方便到达场地,也会限制居民的活动,这种空间私密性强,缺乏与外界交流,造成一种不安全的感觉得,不符合人们在公共场合的基本心理特征。所以利用率会最低。这种模式最不宜营造多功能户外活动空间。

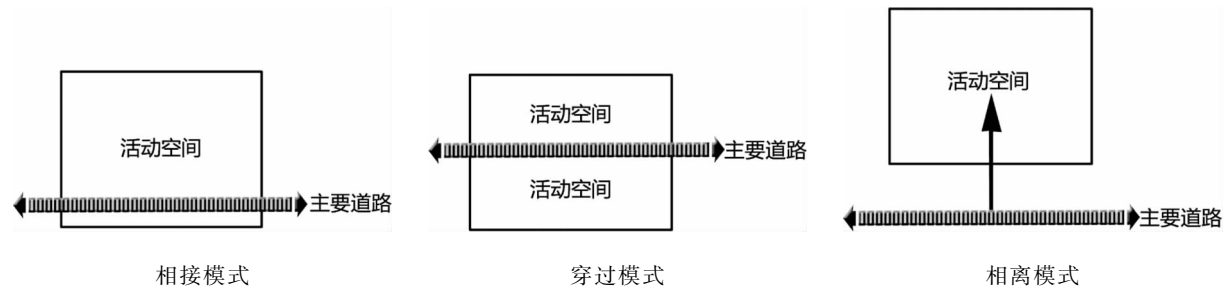


图 5—3 主要道路与场地的关系模式

(图片来源:作者自绘)

通过统计,所有调研的场地中主要道路与场地的关系的数据如下:

表 5—4 调研场地空间主要道路与场地关系统计数据

居住区	主要道路与场地关系		
	穿过模式	相离模式	相接模式
国典华园	0	2	3
华清嘉园	1	0	2
富润家园	0	1	3
华盛家园	0	1	4
锦秋家园	0	1	3
逸成东苑	4	1	3
总计	5	6	18

整理数据发现目前高密度居住区内场地与主要道路的关系中相接模式所占的比例更大,穿过模式与相离模式的道路关系数量相对较少,符合多功能的需求。

遮阴:

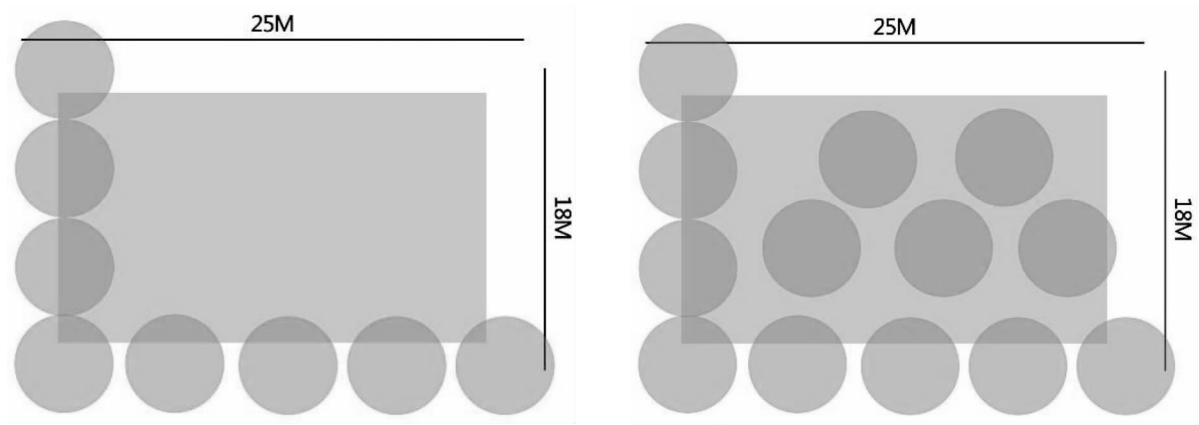


图 5—4 增加场地内部种植,提高遮阴效果

(图片来源:作者自绘)

在场地南侧设置高大乔木,能够形成一定面积的阴影区,但有时可能不能满足尺度稍大的空间。一棵成年乔木冠幅为 5m 左右,因此若保证户外活动场地有良好的遮阴条件,需通过种植分割为不同的遮阴区域,提高遮阴效果。如图,左侧的场地边缘种植并不能满足场地内部对遮阴的需求,因此在右侧图场地中间适当增加种植可以提高场地的遮阴面积,改善活动场地的环境条件。

下表通过对其他变量相似,遮阴条件变化的空间进行整理作出比较如下:

表 5—5 调研场地遮阴条件比较

场地编号	主导活动类型	平均总活动人数	遮阴条件
国 2	照看孩子	36	场地周边的乔灌木
华 1	健身和锻炼	135	场地内与周边的乔灌木
华 3	健身和锻炼	53	场地内与周边的乔灌木
富 1	照看孩子	49	场地周边的乔灌木
富 3	照看孩子	19	场地周边的乔灌木

通过上表可以得出夏季健身和锻炼这项活动对遮阴的需求较高,同时当场地具备内部遮阴时,可以更好地为活动开展提供条件,增加参与活动的人数。

休息设施:

休息设施主要包括一般性座椅(包括棋台等)、亭子廊架等构筑物、功能性休息设施。



一般性座椅



构筑物



功能性休息设施

图 5—5 休息设施的类型

(图片来源:作者自摄)

通过统计,所有调研的场地中休息设施的类型数据如下:

表 5—6 调研场地空间休息设施统计数据

居住区	休息设施类型		
	一般性座椅	构筑物	功能性休息设施
国典华园	13	1	2
华清嘉园	11	1	6
富润家园	10	1	0

居住区	休息设施类型		
	一般性座椅	构筑物	功能性休息设施
华盛家园	7	1	9
锦秋家园	12	0	0
逸成东苑	17	3	12
总计	69	7	29

整理数据发现,目前居住区中的休息设置还停留在以一般性座椅为主的层面上,休息设施的设置变成了为了满足功能的机械性操作,因此一定程度上影响户外活动空间的使用率。精心设计的休息设施在满足基本功能的基础上具有潜在的吸引人流、增强互动、增加活动类型、提高空间利用率的功能,所以户外活动空间建议多采用带有座椅形式的花坛、树池、小品等功能性休息设施。

尺度(场地规模):

表 5—7 国典华园场地面积与利用率关系调研数据

空间编号	场地面积(m ²)	参与人数	利用率			
1 号	85	2/0/0/2	0.024	0	0	0.024
2 号	139	40/21/25/57	0.288	0.151	0.180	0.410
3 号	64	7/0/0/5	0.109	0	0	0.078
4 号	36	13/0/1/1	0.361	0	0.028	0.028
5 号	38	3/1/2/2	0.079	0.026	0.053	0.053

表 5—8 华清嘉园场地面积与利用率关系调研数据

空间编号	场地面积(m ²)	参与人数	利用率				
1 号	500	68/241/90/78/200	0.136	0.482	0.180	0.156	0.400
2 号	525	27/92/10/23/78	0.051	0.175	0.019	0.044	0.149
3 号	113	35/79/18/34/98	0.310	0.699	0.159	0.301	0.867

表 5—9 富润家园场地面积与利用率关系调研数据

空间编号	场地面积(m ²)	参与人数	利用率			
1 号	200	143/23/6/23	0.715	0.115	0.030	0.115
2 号	16	26/11/2/10	1.625	0.687	0.125	0.625
3 号	98	45/18/1/11	0.459	0.184	0.102	0.112
4 号	45	5/8/1/3	0.111	0.178	0.022	0.067

综合以上多个利用率调研表格数据,得出所有场地尺度与利用率的关系如下:

表 5-10 调研场地空间面积与利用率关系统计数据

国典华园		华清嘉园		富润家园	
场地面积(m ²)	利用率	场地面积(m ²)	利用率	场地面积(m ²)	利用率
85	0.015	500	0.164	200	0.244
139	0.268	525	0.072	16	0.766
64	0.061	113	0.367	98	0.214
36	0.129			45	0.095
38	0.053				

(填充格为利用率最高,富润家园2号场地的特殊性除外)

通过对调研的数据分析,从场地面积与场地平均利用率之间的关系得出,高密度居住区户外活动场地中 140m²~200m² 左右的中型空间最受欢迎,利用率最高,空间尺度使人感到亲切,525m² 左右为较宽松的人性化尺度,但不利于高使用率和功能的最大化,因此场地规模小于 525 平方米左右为宜。目前调研的居住区户外活动空间尺度小于 100m² 的占多数,10m×10m 的空间可以布置亭子或廊架,也可以在四周布置休息坐椅,人可以分散成一两个小组活动,这样的空间可以容纳一些老人来参与其中,但限制小孩与青年的活动发生;15m×15m 左右的户外活动场地比较适合设计具有一定多功能性的活动空间,若单纯的设置休息功能会显得有些单调。

(2)不同主导活动类型的场地需求

健身和锻炼:

表 5-11 主导活动类型为健身和锻炼的场地特征总结

		华 1	华 3
空间限定与边界	限定因素	高差	小品
	限定面数	4	4
	限定强度	强	强
主要道路与场地关系	关系	相接	相接
	道路对场地影响	较小	较小
	道路宽度	3m	1.5m
遮阴	绿化与场地关系	位于场地内、周边	位于场地内、周边
	绿化形式	乔木、灌木	乔木、灌木
	遮阴效果	好	好
休息设施	形式	座椅、花池	座椅、廊架、矮墙
	数量	3	8
尺度	具体尺寸	500m ²	110m ²
	认识感受	开敞	围合

它们所具备的共同特征是:

(1)具有明显限定的空间范围。

(2)主要道路与场地关系为相接。

(3)具备场地内部遮阴。

(4)休息设施完备,除了一般性座椅外还有可供休息的其他功能性设施如花池等。

(5)为了保证健身和锻炼的展开,如需要满足一定队形的锻炼或太极、练剑等活动,场地尺度一般偏大且地面平整。

照看孩子:

表 5-12 主导活动类型为照看孩子的场地特征总结

		国 2	国 3	华 2	富 1	富 3
空间限定 与边界	限定因素	铺装、植物	花池、植物	小品、植物	植物、铺装	植物、铺装
	限定面数	2	3	2	4	1
	限定强度	较弱	较强	较弱	较强	弱
主要道路与 场地关系	关系	相接	相离	穿过	相接	相接
	道路对场地影响	中	小	大	较小	较小
	道路宽度	3m	3m	3m	3m	5m
遮阴	绿化与场地关系	位于场地周边	位于场地周边	位于场地外	位于场地周边	位于场地周边
	绿化形式	乔、灌木	乔、灌木	乔木	乔、灌木	乔、灌木
	遮阴效果	较差	较好	较差	较好	较好
休息设施	形式	无	座椅、花池	座椅	座椅	座椅
	数量	0	2	2	4	4
尺度	具体尺寸	139m ²	64m ²	525m ²	200m ²	98m ²
	认识感受	开敞	围合	深远	开敞	开敞

这种空间一般为:

(1)较为开放的限定空间,以两面限定和三面限定为主。

(2)主要道路与场地关系同样为相接。

(3)夏季要具备一定面积的荫凉,冬季要保证一定面积的光照。

(4)无论什么形式的休息设施一定要具备。

(5)为了保证开放性,尺度不能过小,但可以小于以健身和锻炼为主导活动的空间,这样可以保证空间整体的亲切感。

主导活动类型为照看孩子的场地大部分同时也是利用率较高的场所,说明在人流聚集的地方这项活动更易发生。这类空间边缘应尽量丰富,从而方便婴儿车的停靠。

休息和交流:

表 5-13 主导活动类型为休息和交流的场地特征总结

		国 4	富 2
空间限定与边界	限定因素	铺装、植物	廊架
	限定面数	2	4
	限定强度	较弱	强
主要道路与场地关系	关系	相接	相接
	道路对场地影响	中	较小
	道路宽度	3m	3m
遮阴	绿化与场地关系	位于场地周边	位于场地外
	绿化形式	乔木、灌木	灌木
	遮阴效果	好	差
休息设施	形式	座椅	座椅
	数量	5	4
尺度	具体尺寸	36m ²	16m ²
	认识感受	局促	封闭

这类空间的特点为:

- (1)不需要很大,为了避免外界干扰基本上私密性较强。
- (2)主要道路与场地关系最好避免穿过模式。
- (3)休息设施是必要因素。
- (4)如果光照条件良好能促进此类活动的发生。

设计时不会专门考虑以休息和交流为主的空间,一般和其他活动空间相结合来设置,但调研结果发现产生了这种空间,说明这类空间过于单调,只具备一般性座椅同时受空间尺度限制其他活动无法展开。

其他:

表 5-14 主导活动类型为其他的场地特征总结

		国 1	国 5	富 4
空间限定与边界	限定因素	花池、座椅、植物	铺装、植物	植物
	限定面数	3	2	2
	限定强度	较强	较弱	较弱
主要道路与场地关系	关系	相离	相接	穿过
	道路对场地影响	小	中	大
	道路宽度	3m	3m	1.5m
遮阴	绿化与场地关系	位于场地周边	位于场地周边	位于场地周边
	绿化形式	乔、灌木	乔、灌木	灌木
	遮阴效果	较好	好	较差
休息设施	形式	座椅、花池	座椅	座椅
	数量	3	5	3
尺度	具体尺寸	85m ²	38m ²	45m ²
	认识感受	围合	局促	围合

这类空间发生的活动类型包括站立张望、打电话、看书、遛狗等。这类空间无明显相似特征,活动发生的随机性大,与休息和交流为主导的空间相似,以其他活动类型为主的空间环境条件一般,例如国典华园1号场地中乒乓球台占据主要空间,5号场地为只具备座椅的小尺度空间,富润家园4号场地尺度小同时受穿过人流与建筑阴影影响干扰大,此种活动类型对空间的要求最低。因此满足其他活动类型的空间即可满足此类活动的发生。

参 考 文 献

专著:

- [1] 李道增. 环境行为学概论[M]. 北京:清华大学出版社,1999
- [2] 白德懋. 居住区规划与环境设计[M]. 北京:中国建筑工业出版社,1993
- [3] 杨·盖尔. 交往与空间(何人可译)[M]. 北京:中国建筑工业出版社,1992
- [4] 芦原义信. 外部空间设计[M]. 北京:中国建筑工业出版社,1985
- [5] (美)克莱尔·库珀·马库斯,(美)卡罗琳·弗朗西斯编著,俞孔坚等译. 人性场所[M]. 中国建筑工业出版社,2001
- [6] 荆其敏,张丽安著. 城市休闲空间规划设计[M]. 东南大学出版社,2001
- [7] 梁雪,肖连望编著. 城市空间设计[M]. 天津大学出版社,2000

学位论文:

- [1] 杨子夜. 北京市居住区户外环境与居民行为模式的关系研究[D]. 北京林业大学,2010
- [2] 武婵燕. 居住区交往空间环境设计研究[D]. 浙江大学,2005
- [3] 郭珊珊. 中高强度住区户外公共活动场地空间设计研究——以成都市为例[D]. 西南交通大学,2011
- [4] 于冬亮. 住区交往空间研究[D]. 同济大学,2007
- [5] 曹悦. 基于环境行为学的大学校园户外空间研究[D]. 北京林业大学,2010
- [6] 李夏艳. 居住区绿地设计中行为与心理研究[D]. 中南林业科技大学,2007
- [7] 马岩. 现代城市街道环境中的小型休憩绿地设计探讨[D]. 北京林业大学,2004
- [8] 王昊. 城市道路景观形态研究[D]. 东南大学,2006
- [9] 陈平. 论社区景观环境的私密性设计[D]. 天津美术学院,2011
- [10] 孙景. 城市公共场地的多功能性[D]. 北京林业大学,2005
- [11] 韩秀茹. 城市公园中休闲空间研究[D]. 北京林业大学,2006
- [12] 陈笑. 居住小区加强邻里交往的园林环境设计浅析[D]. 北京林业大学,2010
- [13] 曾锐. 现代高层住区的交往空间研究[D]. 湖南大学,2011
- [14] 陈鸿. 城市居住小区邻里交往空间研究[D]. 西南交通大学,2003
- [15] 姚中. 邻里交往空间的有效性[D]. 同济大学,2008
- [16] 朱瑞冉. 居住区中心绿地开放性设计研究[D]. 哈尔滨工业大学,2008
- [17] 韩立波. 高速公路服务区景观规划设计理论初探[D]. 南京林业大学,2003
- [18] 陈平. 论社区景观环境的私密性设计[D]. 天津美术学院,2011
- [19] 王颖辉. 西安市住区宅间交往空间规划设计研究[D]. 西安建筑科技大学,2007

期刊文章:

- [1] 林文洁,于喆,杨绪波. 居住区老年人户外活动及其空间特征研究——夏季户外活动实态调查[J]. 建筑学报,2011(2): 73—77
- [2] 张玲玲. 环境心理学中的微观空间行为[J]. 黑龙江科技信息,2009

- [3] 宋鸣笛,蔡强. 关于现代住区活动场地及设施的设置的思考[J]. 城市环境设计,2005(3)
- [4] 海伦伍勒,于一平. 人性化公共开放空间[J]. 世界建筑,2006(7)
- [5] 章俊华. 环境设计的趋势——“公众参与”[J]. 中国园林,2000(1)
- [6] 吴卉,董雅. 城市公共开放空间设计与环境行为研究[J]. 建筑学报,2012(S2)
- [7] Deqing Bu, Zhao Wang. Research on Spatial Identities of Activity Spaces in Urban Parks Neighboring Residential Areas. Applied Mechanics and Materials [C]. Switzerland: Trans Tech Publishers,2014
- [8] 中华人民共和国建设部. 城市居住区规划设计规范[J]. 中国建筑工业出版社,2002

Research on utilization of open grounds and high density residential multi functions in Beijing

Wang Zhao

(School of Architecture and Art, North China Univ. of Tech. , Beijing, 100144)

Abstract In recent years, Beijing region showing growing population density while the status of the land tends to be limited. The key thesis of this paper is how to solve the question the needs of space and function in a residential area reasonably under the high-density development mode, or how to better reach both the balance between the needs of space and function. This paper takes some of open grounds in such typical urban residential quarters as GuoDianHuaYuan, HuaQingJiaYuan, and FuRunJiaYuan in Beijing as object of study. At-site survey was conducted based on theories and methods of environmental behavior studies, involving classification of users and activities, and detailed recording of user groups' behaviors by means of counting and photo taking. On basis of such, functions and utilization of open grounds were summarized for in-depth analysis of the relation between the enclosing, shading, rest facilities, scale, location and frequency of activities taken place. And the influence of such relation on utilization of open grounds is concluded, which is expected to guide the designing of open grounds in residential quarters.

The first chapter describes in the the background of crowed population and limited land of Beijing, the urgency is to increase high-density residential outdoor activity space utilization. Currently on the site and space research are part of the residential area of that site requires multi-functional proposal, there is no specific article describes how to multipurpose venue or space research, innovation of this paper is to propose concrete design suggestions.

The second chapter of this paper is define the concept of “high-density residential area”, “outdoor space utilization,” and “outdoor space”, space for outdoor activities. Besides, the elaboration of constituent elements are described and the theory of exterior space design are introduced.

The third chapter elaborate the characteristic of the residential space for outdoor activities is restricted, publicity and territoriality. The manifestation mode of versatility outdoor space is by using the crowd, according to time and by type of activity. The type of activity appears in residential areas to classify to facilitate analysis of the later chapters.

The fourth chapter is the key sections of this paper. Research and analysis of Beijing typical high-density residential area of space for outdoor activities is mainly included. Through the high-density residential area of the overall design of outdoor activity space, different aspects of the intended function, staff activities and changes in the characteristics of spatial capabilities to statistical analysis, to give the corresponding research data, analysis of specific factors that may have an impact on the function, summed up relatively high utilization venues possess characteristics.

The fifth chapter is also focused on sections of this article, the fourth chapter in the foundation of research may affect the function of the factors of space, depth of enclosed living area, shade, rest facilities, scale, location and so on in the creation of multi-functional space for outdoor activities during the impact. Summed up the different requirements of different types of activities on the site, make the proper function integration, division of space level, increasing details of the deal, versatile control and other design elements meanwhile.

Chapter VI of the paper concludes that both sides enclosed and enclosed on three sides theoretically should be the main enclosure multipurpose venue way; the relationship between the main road and site recommendations for the “space integration”; about 140 square meters — 200 square meters medium-sized multi-functional space recommend in this article, etc.

Key words high density resident area; open ground; multi-function; research

张家口近代城市规划建设研究

研究生 范家昱

导师 于海漪

(北方工业大学建筑与艺术学院 北京 100144)

摘 要 张家口是京津冀地区重要城市之一,在近代前主要担负军事防御的职能。随着京张铁路的建设与张家口开埠,成为张家口近代城市建设的起点。本文从城市规划学的角度出发,运用文献调查法、现场调查法与比较分析法,对张家口近代 1902—1952 年间的城市建设实践进行梳理,对并其城市地位与城市职能、空间形态、城市建设机构与组织等方面进行综合分析,从而发现近代城市建设与城市在区域中的地位密不可分,以及多主体协同规划对城市发展的重要作用。

本文的研究内容主要有六章共三部分:

第一部分:第一章绪论,论述研究背景、研究内容、国内外研究现状及发展趋势、研究方法与技术路线,以及文研究范围的界定。

第二部分:第二至四章,是本文的重点章节。第二章简述了张家口近代城市建设背景,并以此为基础提出张家口近代城市建设的历史分期。第三、四章,分别从时代背景、主要城市建设实践、建设主体与相关文件等方面,论述了京张铁路促进下的城市建设(1902—1929 年)、动荡的社会环境阻碍城市建设(1929—1948 年)、建国初期恢复城市建设(1948—1952 年)三个历史时期,张家口城市建设与发展的主要内容与特点。

第三部分:论文第五章,对张家口近代城市规划建设的特点进行分析,总结其在城市地位与城市功能演变、城市职能、空间形态、建设主体等方面的特点,对京津冀协同发展提供借鉴。

本文旨在梳理张家口近代的城市规划建设史,以京津冀区域为视角,研究在京张铁路的促进下张家口由军事堡垒逐渐演变为趋于区域复合中心的发展特点和演变规律,总结历史经验,为京津冀协同规划与研究、促进京津冀协同发展提供基础资料。

关键词 京津冀;张家口;近代城市规划与建设;非政府组织

作者简介 范家昱,女,1989 年出生,硕士研究生,主要研究方向为城乡规划历史与理论。

E-mail:497573960@qq.com

1 引言

1.1 研究背景

张家口市是京津冀地区的重要城市之一,梳理研究张家口近代的城市规划建设史不仅可以总结历史经验,分析不同因素对张家口的城市形态、经济与城市地位等的影响,总结历史的经验,为京津冀地区的研究提供借鉴,以促进华北区域内陆城市的共同发展。

1.2 研究内容

本课题以张家口近代城市建设史为研究对象,以传统到近代的历史发展为纵向分析轴,以其与华北区域的联系为横向分析轴,主要针对城市建设、交通及其与华北区域的关系进行分析研究,整理张家口近代城市建设史。在厘清上述历史的基础上,通过整理张家口的交通、经济等基础条件,分析张家口在华北区域中的交通枢纽地位,分析张家口与华北区域关系的演变过程。

1.3 研究范围界定

1902年,中俄签订《依露支条约》,成为张家口近代城市发展的开端。1952年,察哈尔省撤销,河北省建制,张家口的城市发展进入较为平稳的状态。故本文中关于张家口近代城市建设的时间段界定为1902—1952年。

目前,张家口是河北省下辖地级市,现市辖13个县,总面积3.68平方公里。在1902—1952年这一期间内,张家的主要建设实践集中在现市区范围内,即桥西区与桥东区,这一范围也可看作是整个张家口发展的缩影,其城市职能、特点集中体现于此。故本文将近代张家口的范围界定为桥东区与桥西区(见图1—3)。

2 近代张家口城市建设背景与历史分期

本章从张家口近代城市建设背景入手,简要阐述张家口地理、历史沿革以及其城市发展的经济和政治背景,以不同时期张家口在区域中承担的不同职能为线索,结合时代背景,提出张家口近代城市建设的历史分期,并简述各时期城市建设主要内容。

2.1 张家口近代城市建设背景

通过对张家口近一个世纪的城市建设沿革的简要梳理,作者认为近代张家口的崛起与其在区域中所承担的职能密不可分,同时奠定了其在京津冀区域中的重要地位。今天,张家口以其优越的地理位置、独特的环境优势,在京津冀协同规划中扮演着重要的角色。

2.2 张家口近代城市发展基础

张家口形成之初是为了军事防御,从用地比例、人口比例等方面分析,张家口是一座不折不扣的武城。而随着商业、对外贸易的出现和不断发展,张家口成为华北地区著名的商埠,并带动了周边地区的发展。

张家口在1902年前的城市格局已经初步显现,主要体现在城市功能分区以及城市结构的出现,上、下两堡一南一北,成为张家口最初城市功能分区的雏形随着张家口不断进行近代城市建设,其城市地位逐渐提高,城市空间开始向桥东区拓展。

2.3 张家口近代城市建设历史分期

本节以张家口近代城市的发展脉络与特点为主要依据,将张家口近代城市建设划分为以下三个历史时期。即京张铁路促进下的城市建设(1902—1929年)、社会动荡环境制约城市建设(1929—1948年)、建国初期恢复城市建设(1948—1952年)。

2.4 本章小结

本章通过阐述了张家口得天独厚的自然、地理环境与历史沿革,阐述了张家口近代城市建设的背景,为张家口近代城市发展奠定了坚实的基础。通过研究张家口近代城市发展脉络和建设特点,提出张家口近代城市建设的三个历史分期,并分别对其特点及主要建设内容进行阐述,发现张家口近代城市的崛起、发展与其不同时期在京津冀区域中所扮演的角色有着密不可分的联系。

3 京张铁路促进下的城市建设(1902—1929年)

本章研究的时间范围是1902年张家口局部开放为通商口岸至1929年。1902年以前的张家口建成区主要集中在桥西区,且城市功能较为单一。在京张铁路等有利因素的促进下,经过二十余年的建

设与发展,张家口迎来了近代城市建设的大发展,涌现出一大批机构和组织参与城市建设,形成了商业街区、铁路附属设施,推动了城市的全面发展,内容涵盖商业、工业、公共设施等方面的建设,城市功能日趋完善,城市地位逐渐提高。

3.1 时代背景

京张铁路与张家口火车站的修建、开埠、随即成为察哈尔特别区的首府,为张家口近代城市建设与发展带来强有力的推动作用,各方面因素都为张家口在这一时期城市商业的飞速发展创造出有利条件。

3.2 京张铁路及其他交通设施的建设

这一时期,张家口的交通设施飞速发展,内容涉及铁路、公路及相关设施,为张家口城市各方面的发展提供契机,形成张家口地区与察哈尔部近代铁路网。

3.3 掀起张家口近代城市建设的高潮

张家口这一时期的城市商业能都得到迅速发展,在张家口火车站附近开辟街道,兴建商号住宅,吸引商户,创建商业街,并结合已有的商业街,建设娱乐场所,吸引客流,催生多种商业形式。

城市的近代化离不开工业,张家口各种官办或商办的工业陆续兴起,这些近代企业的创办,不但使得张家口从过去单一的消费性结构转向生产性结构,也使张家口这个历史悠久的商埠向工业城市迈进,加快了张家口城市近代化的脚步。

随着京张铁路的开通,铁路职工人数越来越多。解决与日俱增的铁路工人日常生活问题,怡安公司出资兴建居住区,即马路街,街区面积约为 100 公顷。

张家口市建设了市政工程,例如市内交通设施、通信设施、桥东区西村及其他公共服务设施。

3.4 城市建设主体与相关文件

通过对这一时期张家口城市建设的梳理,发现怡安公司以非政府组织的身份参与城市建设,并促进了张家口城市格局的形成,是近代城市规划思想的具体体现;而冯玉祥由宗教信仰和借鉴成功经验作为城市建设的思想来源,对张家口进行城市建设,带有区域性、实践性、可持续性的特点。非政府组织参与城市建设,为张家口的近代城市发展做出巨大贡献。

张家口商埠局作为察哈尔省政府下设机构,主理城市建设、市政管理等方面的工作,并出台《市政工程七项计划》,是张家口最早的关于市政工程的文件,标志着张家口的市政工程和基础设施建设逐步走向规范化、制度化。

3.5 1902—1929 年张家口城市建设特点

在 1902—1929 年这短短的二十年间中,受到京张铁路的开通的有利推动,张家口成为华北地区重要的交通、贸易枢纽,更是为张家口的政治、经济等各方面起到促进作用,并对张家口近代城市发展提供了良好的契机,主要体现在城市空间格局的变化、城市地位提升和城市近代化水平逐渐提高这三方面。

3.6 本章小结

由于京张铁路的建设和张家口开埠等众多有力因素的促进,掀起了张家口近代城市建设的高潮,促进了张家口近代城市的大发展。同时涌现出一大批具有近代规划思想的机构或组织,经过短短二十余年的建设,为张家口近代城市建设做出巨大贡献,张家口一跃成为区域性的治、交通、商业复合中心。

4 城市建设的阻碍与恢复(1929—1952 年)

本章研究的时间段为 1929 年中俄断交直至 1952 年新中国成立初期。自 1929 年起张家口处于内

忧外患的时期,军阀混战、战争不断、中俄断交、自然灾害等,这样的社会背景直接导致以对俄贸易为主导的张家口经济陷入困境。而 1931 年爆发的九一八事变开始,接连战争迫使张家口陷入经济萧条,城市管理混乱的境况。1948 年人民解放军收复张家口,张家口社会运转逐渐恢复其战前水平,并在共产党的领导下,逐渐完成从商业城市向工业城市的转化。

4.1 时代背景

中俄断交、抗日战争与解放战争的爆发,动荡的社会环境使张家口的城市建设受阻。此后 1948 年解放战争胜利直至 1952 年“一五”计划开局之年,张家口逐渐恢复城市建设,并成功完成商业城市向工业城市的转型。

4.2 动荡的社会环境阻碍城市建设(1928—1948 年)

日伪统治时期的城市建设主要以工业设施和城市公共设施为主,目的是获取最大利益,并且抢占民宅,实行粗暴的都市计划,延缓了张家口市的近代化进程。解放战争时期,在张家口的城市建设主要以市政公用设施为主,目的是为了恢复张家口的城市运转。

4.3 建国初期恢复城市建设(1948—1952 年)

1948 年底,人民解放军收复张家口,张家口市建筑局负责张家口市城市建筑管理事宜,建立稳定的社会秩序,将工业、商业、农业的恢复置于重要地位。1953 展开“一五”计划,充分发挥张家口的资源特色,逐渐转变为工业城市。

4.4 城市管理职能机构及相关文件

由于社会环境动荡不安,1929 年至 1952 年间,张家口城市建设先后受察哈尔省建设厅、都市建设处、张家口市人民政府、国民党张垣市政府的管理。张家口市人民政府在《关于入城工作的指示》的基础上,细化了保护条目,依据张家口特点,颁布了《入城条例》,对张家口市已有的各种城市公共设施予以保护。

4.5 本章小结

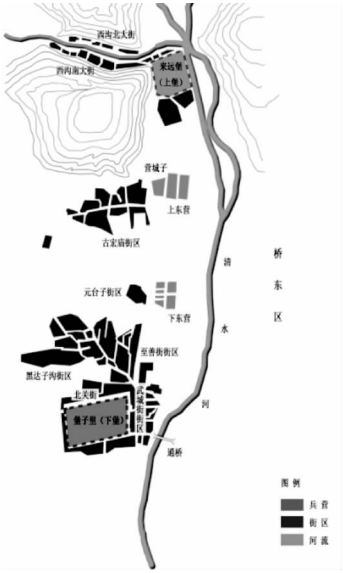
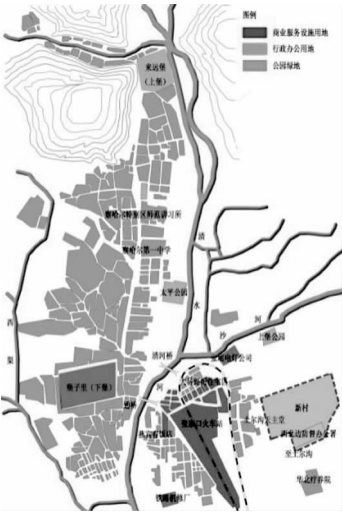

这一阶段,张家口先后经历了城市发展的阻碍与恢复。经过接连的战争,张家口的城市发展一度陷入停滞。日伪政府统治期间,为了谋求利益的最大化,侵占了张家口大片城市用地,强占市民房屋,建设了带有侵略殖民色彩的公共设施和工业设施。解放战争时期,政权频繁变换,张家口最终迎来解放。经过战争,张家口的城市空间没有明显的扩张。在建国初期,张家口桥东区的中心区域仍处于火车站附近。失去了对邻国经济依赖性较强的对外贸易,反而促成了张家口从先前的区域性复合中心向现代工业城市的转型。在中国共产党的领导下,建国初期以恢复城市工业生产为首要任务,在张家口兴建了一批工业设施,促进了张家口的工业蓬勃发展。

5 张家口近代城市规划建设的特点

5.1 张家口城市地位与城市功能的演变(表 1)

纵观张家口近代从城市形态、职能的演变,与其在华北区域内所承担的功能密不可分,导致在各个不同阶段,城市建设的特点也不尽相同。

表 1 近代张家口城市功能与地位演变

	开埠前的张家口	1902—1929 年	1929—1952 年
发展进程	对外贸易初期	城市高速发展	阻碍与恢复
区域职能	对蒙、俄商贸	政治、交通、物流、经济、金融复合中心	华北重要战场、工业城市
城市空间			
建设特点	除堡子里、来远堡和兵营外，大部分为自发形成的街区，为对外贸易服务，并催生一系列依靠对外贸易的商业形式。	城市近代化的兴起，建设城市公共设施，注重城市防灾工程的兴建，经济中心向桥东区拓展，从而形成桥西、桥东二区传统与现代城区相结合的特点。	城市空间依旧延续上一时期的空间形态，城市建设以市政工程为主，并逐步将城市空间拓展至桥东区南部。

5.2 城市职能的转变

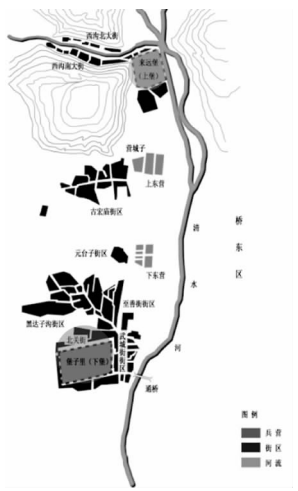


张家口的商业从明朝就开始逐渐兴起，历史悠久。近代时期，张家口在各个阶段创办了不同类型的商业。随着时间的推移，张家口的商业受到多种因素的共同影响，失去了区域商业中心的职能，商业发展日趋平缓。

张家口的城市近代化另一大特点就是对工业的重视。不论是经济繁荣时期建设的民族资本主义工厂，还是在建国初期为了恢复工业生产建立的皮毛加工厂，张家口的工业一直被看作是城市发展的重中之重。各个时期，张家口兴建的工业类型各有侧重，建设目的也随着城市发展进程的推进而变化。

5.3 城市空间形态的演变(表 2)

张家口近代的城市中心几经发展演变，随着张家口在区域中承担的职能不同，加之各种影响因素先后产生，其城市中心区也随之在数量、类型等方面产生变化。

表 2 近代张家口城市空间形态的演变

		开埠前的张家口		1902—1929 年			1929—1952 年	
近代城市发展进程		对外贸易初期		高速发展			平稳与恢复期	
区域定位		对蒙俄的商贸及物流集散地		政、交通、经济复合中心			华北重要战场、工业城市	
城市中心	数量	2 个		3 个			2 个	
	类型	军事中心	贸易中心	政治中心、贸易复合中心	传统商业中心	交通中心、近代商业中心	传统商业中心	近代商业中心
	区位	堡子里	来远堡	来远堡	堡子里	火车站周边	堡子里	火车站周边
	影响因素	军事防御	规范对外贸易	察哈尔省首府	军事职能淡退	近代交通的兴起	恢复并延续战前城市中心	
	区位							

张家口近代的城市道路结构可以反映出张家口城市演进的过程。在城堡周边的道路根据城堡城墙的走向呈现出方格网式的道路结构。城堡外的道路多是居民自发形成的,道路多为不规则结构。随着时间的推进,在近代城市规划思想的影响下,火车站附近的道路结构依然呈现出了规则的、方格网式的道路结构,并延续至今。

5.4 多主体共同参与城市建设

政府机构和非政府组织共同参与了张家口近代的城市建设。据统计,参与到张家口近代城市建设的官方机构或非官方组织共计 19 个之多。官方机构从宏观角度出发,为张家口城市发展提供引导,非官方组织作为城市的利益相关者和参与者,直接参与城市的规划建设,这种协同规划的城市建设模式,使得张家口近代的城市发展突飞猛进。

5.5 京津冀区域中近代本土城市规划实践的典型案例

张家口近代崛起的过程,说明它的肇始就与其在京津冀区域中的区位有极大的关系,近代时期张家口在京津冀区域发展中的作用,对当前京津冀协同发展有积极的启发作用。张家口城市的近代化与其对外贸易的繁盛息息相关,经济空前繁盛为张家口城市近代化带来了发展的契机。此外,多主体共同参与张家口近代城市规划建设促进了张家口近代城市的发展。近代张家口城市建设实践的历史告诉我们,大量普通城市是从城市自身的历史、地理条件出发,基于当时的技术与社会发展需要,通过本土技术人员,与本地官方、民间等各种组织机构或个人的共同努力,逐步发展起来的,不应忽略本土

的努力和探索。

6 结语

1902年—1952年间,依靠其得天独厚的地位区位优势,以及对外贸易的发展和兴起,张家口从一个军事城堡迅速发展成中外闻名的华北商埠,奠定其在京津冀地区中的重要地位。

通过论文的研究,近代历史表明张家口近代城市建设首先得益于其在区域中的城市地位,说明一个城市在区域中扮演复合自身特点的地位,能够促使城市职能与空间形态与区域相协调,进而推动城市发展。其次,中国近代历史上社会动荡,政府往往无力成立专门机构负责城市建设和管理,多主体、多渠道的机构、组织和个人共同参与城市建设,说明在动员各方利益群体共同参与城市建设时,城市建设呈现出繁荣景象。最后,张家口的近代城市建设奠定了其在京津冀地区中的重要地位,影响了今天和未来的城市发展。

参 考 文 献

专著、论文集

- [1] 张家口市建设档案馆,王林,刘信. 张家口市城乡建设纪事[M]. 中国档案出版社,1998
- [2] 中国人民政治协商会议张家口市委员会文史资料委员会. 张家口文史资料第二十一辑(社会卷·上)[M]. 张家口日报社,1992
- [3] 中国人民政治协商会议张家口市委员会文史资料委员会. 张家口文史资料第二十二辑(社会卷·下)[M]. 张家口日报社,1992
- [4] 张家口市政协文史资料委员会,中共张家口市委党史研究室. 张家口文史资料第二十三辑(察哈尔纪事特辑)[M]. 张家口日报社,1993
- [5] 徐纯性. 河北城市发展史[M]. 河北教育出版社,1991
- [6] 刘振英. 张家口历史文化丛书——兴盛的古商道[M]. 北京:党建读物出版社,2006
- [7] 路联速修,任守恭纂:《万全县志》,张家口概述·官署,民国二十三年(1934年)
- [8] 孙世芳修,乐尚约纂. 宣府镇志? 城堡考(卷11)[M]. 台北:成文出版社,1970:94
- [9] 吴良镛等著. 京津冀地区城乡空间发展规划研究二期报告[M]. 清华大学出版社,2006

学位论文、报告

- [1] 闫浩. “隆庆和议”史料研究[D]. 内蒙古大学,2008
- [2] 刘德勇. 清至民国张库交通与张家口城市商贸发展[D]. 中国社会科学院研究生院,2011
- [3] 段海龙. 京绥铁路研究(1905—1937)[D]. 内蒙古师范大学,2011
- [4] 付丽娜. 察哈尔地区的商业与城市近代化(1840—1935)[D]. 内蒙古大学,2008
- [5] 张志国. 河北省城镇空间布局与发展研究[D]. 河北师范大学,2009
- [6] 黄红海. 近代河北城市功能变迁研究[D]. 中原工学院,2012
- [7] 李晨晖. 明清时期张家口地区商业地理研究[D]. 西北师范大学,2011
- [8] 董花. 明清时期张家口商贸兴衰研究[D]. 广西师范大学,2014
- [9] 梅兰. 近代张家口城市发展研究(1860—1937)[D]. 河北大学,2013
- [10] 闫志弘. 近代张家口皮毛贸易述论[D]. 东北师范大学,2011

期刊文章

- [1] 杨润平. 张家口开埠与城市近代化的起步——1902—1952年历史的回顾[J]. 城市史研究, 2002: 475—488
- [2] 张轶欣. 张家口商业兴衰与近代城市空间的演变[J]. 河北北方学院学报, 2008, 02: 59—63
- [3] 王洪波, 韩光辉. 从军事城堡到塞北都会——1429—1929年张家口城市性质的嬗变[J]. 经济地理, 2013, 05: 72—76
- [4] 高春平. 张库商道之兴衰[J]. 中国名城, 2009, 05: 24—27
- [5] 牛国祯, 梁学诚. 张库商道及旅蒙商述略[J]. 河北大学学报(哲学社会科学版), 1988, 02: 6—11
- [6] 王洪波, 韩光辉. 从军事城堡到塞北都会——1429—1929年张家口城市性质的嬗变[J]. 经济地理, 2013, 05: 72—76
- [7] 汪寿松. 华北自开商埠城市述论[J]. 城市史研究, 2004, 00: 141—154
- [8] 樊如森. 近代华北经济地理格局的演变[J]. 史学月刊, 2010, 09: 87—96
- [9] 王向英, 张常勇. 冯玉祥与察哈尔抗日[J]. 内蒙古社会科学(汉文版), 2004, 02: 27—30
- [10] 吴晓军. 公众参与城市规划的角色和作用[J]. 天府新论, 2011, 06: 98—100
- [11] 何一民, 付娟. 从军城到商城: 清代边境军事城市功能的转变——以腾冲、张家口为例[J]. 史学集刊, 2014, 06: 16—24
- [12] 贾熟村. 对察哈尔抗战的考察[J]. 内蒙古师范大学学报(哲学社会科学版), 2012, 05: 105—108
- [13] 王向英, 张常勇. 冯玉祥与察哈尔抗日[J]. 内蒙古社会科学(汉文版), 2004, 02: 27—30
- [14] 裴赞芬. 近代河北城市化试论[J]. 河北师范大学学报(哲学社会科学版), 1998, 04: 110—115
- [15] 刘海岩. 近代华北交通的演变与区域城市重构(1860—1937)[J]. 城市史研究, 2002, 00: 24—48
- [16] 郝庆合, 殷毅. 京张铁路与天津近代物流[J]. 北京交通大学学报(社会科学版), 2009, 02: 32—37+53

A Study on the History of Modern Urban Planning and Construction of Zhangjiakou City

Fan Jiayu

(School of Architecture and Art, North China Univ. of Tech., Beijing, 100144)

Abstract Zhangjiakou is one of the most important cities of the Capital Region. In modern times, Zhangjiakou City was mainly responsible for the functions of the former military defense. With the construction of Beijing-Zhangjiakou railway, it became the starting point of Zhangjiakou Modern Urban Planning and Construction. From the perspective of urban planning, use literature survey, site investigation and comparative to analyze Zhangjiakou Modern Urban Construction Practice from 1902 to 1952, to sort out city status and urban functions, spatial morphology, urban construction a comprehensive analysis of institutions and other organizations, which found in modern urban construction and urban status in the region are inseparable, as well as multi-agent collaborative planning the important role of urban development.

This thesis is divided into three parts and consisted of total mainly seven chapters.

The first part is the research background, research content, research status and development trends, research methods and technical route, and text scoping study.

The second part is Chapters 2—4, is the key sections of this article. The second chapter outlines the background of Zhangjiakou Modern Urban Construction, made history as the basis of modern urban construction stages of Zhangjiakou. Two chapters, namely in terms of historical background, major urban construction practice, the main building and the relevant documents, discussed the promotion of urban development under the Beijing-Zhangjiakou Railway (1902—1929), the turbulent social environment hinder urban development (1929—1948), the early days of rehabilitation of urban construction (1948—1952) of the three historical periods, the main contents and characteristics of Zhangjiakou city construction and development.

The third part is the fifth chapter, the characteristics of modern urban planning and construction of Zhangjiakou analyze, summarize the evolution of the status of the city and urban functions, characteristics of urban functions, spatial form, the main building and other aspects of co-development of Beijing, Tianjin reference.

This thesis aims to sort out the history of modern urban planning and construction in Zhangjiakou to Beijing, Tianjin and the perspective of research in the Beijing-Zhangjiakou Railway promote Zhangjiakou evolved into a fortress by the military tends to develop the characteristics and evolution of the regional composite centers, summing up the historical experience, as Beijing, Tianjin and collaborative planning and research, and promote coordinated development of Beijing, Tianjin and biography.

Key words he capital region; zhangjiakou city; early modern city planning and construction; non-governmental organization

基于公共健康需求的公共空间适老性研究

研究生 曹 阳

导师 梁玮男

(北方工业大学建筑与艺术学院 北京 100144)

摘 要 以往对于公共空间适老性的研究大多基于老年人的生理及心理特征,本文从全新视点即老年人公共健康需求角度对公共空间进行研究。国外对于公共健康与城市设计的关系的研究主要从景观设计、城市设计两方面入手,公共健康与绿地景观的研究著作及在各国实践项目逐渐增多,美国于2010年制定纲领性文件:“积极设计导则”(Active the Pedestrian Space Design Guidelines),其中提出分析城市设计与积极的运动方式的关键是“5D”变量。基于“5D”变量,针对我国老年人这一特定群体的健康需求及公共空间需求,提出公共空间适老性的新“5D”变量即密度(density)、多样性(diversity)、步行空间设计(the Pedestrian Space Design)、目的地可达性(destination accessibility)、容错性(damage reduction),进而探讨相关设计原则与策略,并针对案例社区提出改进建议。

关 键 词 公共健康;公共空间;适老性;北京

作者简介 曹阳,女,1988年出生,硕士研究生,主要研究方向为建筑设计及其理论。

E-mail:815868095@qq.com

1 绪论

1.1 研究目的与意义

基于我国老龄化社会国情,本文试图探寻老年人公共健康需求与社区公共空间设计的关系,研究引导老年人积极健康的生活方式的社区公共空间设计策略。研究意义在于:一方面,可以为城市公共空间设计提供一定参考,使之满足并适应老龄化的发展趋势与要求;另一方面,基于公共健康需求的公共空间研究尚属崭新的课题,可以在一定程度上拓展公共空间设计研究,因此本文的研究具有理论及实践意义。

1.2 国内外研究综述

对于公共健康与城市设计关系的研究,美国于2010年制定“积极设计导则——让设计促进市民运动和引导健康生活方式”(Active the Pedestrian Space Design Guidelines: promoting physical Activity and Health in the Pedestrian Space Design),导则是第一篇关注设计师在市民公共健康问题上的重要地位的出版物。在景观设计学科中,克莱尔·库珀·马库斯于二十世纪九十年代中期开始进行的治疗性花园研究,将园艺实践与健康实践相结合,强调人工设计景观在患者康复和治疗中的作用性,进入二十一世纪后,关于公共健康与绿地景观的研究著作继续增多。

2013年李煜、朱文一在《世界建筑》上发表“纽约城市公共健康空间设计导则及其对北京的启示”文章中,首次提出中国城市公共空间设计应基于公共健康需求这一理念。同年以俞孔坚教授为首,在景观设计领域探讨公共健康与绿地景观设计的关联,阐述了国外一些研究如城市绿色景观对大众健康的影响机制及重要研究问题。

对于社区老年人公共空间的研究,发达国家由于比中国早进入老龄化社会,因此对老年建筑的研究已颇具成效,理论上以关于老年建筑的规范标准和场地设计为主;实践上侧重对老年公寓、老年住宅的探索。

在理论研究及相关设计标准方面,我国已出版一些针对老年人的设计规范和标准,同时近年来,国内学者对老年建筑与老年人行为及心理特征、老年人社区公共空间、养老设施等研究日益增多,涌现出来一批较为优秀的硕士论文和博士论文。

1.3 相关概念阐释

① 公共健康

从城市设计层面上讲,公共健康问题主要指城市化带来的负面影响对“公众”即全体市民生理及心理健康的威胁。城市已建成环境过多地鼓励了消极的生活方式如看电视、乘坐私家车、吃快餐等,使市民过多依赖外界能源而过多消耗自身热量,从而导致肥胖及其相关疾病,及抑郁、焦虑、紧张等心理疾病,与此同时,水质污染、土壤污染、噪声污染,空气污染等环境问题也相继衍生出来,对城市环境造成负担进而影响公共健康。

② 公共空间

本文主要研究的是如何设计公共空间使其更适宜老年人活动与交往,而老年人由于受到生理因素的限制及心理因素的影响,室外活动主要发生在一定区域范围即社区内,到离家较远的市域公共空间频率较少,因此本文研究的范围缩小,限定在社区公共空间中。

笔者认为,“公共空间”与“私人空间”对应,强调的是“公共”这一概念。从这一角度来看,研究城市公共空间,不仅应研究城市中供市民自由使用的外部公共空间,也应研究建筑内部公共空间如楼梯间、电梯间等。

③ 适老性设计

本文所研究的“适老性”,研究对象主要指在城市普通社区的公共空间层面,满足老年人健康需求,如生理、心理、行为等各个方面所对应的功能设施的综合性能。参照《老年人居住建筑设计标准》、《老年人建筑设计规范》等相关规范,本文探讨的“适老性”包含如下几方面内容:容错性、适应性、安全性、可达性、易辨性等。

1.4 研究方法

研究方法为:文献分析法、归纳与演绎的逻辑方法、理论联系实际、系统研究方法。

2 公共空间适老性设计之新“5D”变量构建

2.1 积极设计导则——公共健康与城市设计研究的纲领性文件

“5D”变量(The Five “D” Variables)在美国于2010年制定《积极设计导则——让设计促进市民运动和引导健康生活方式》导则中被提出。过去15年的超过150个研究,都在探讨城市设计与积极的运动方式如步行、跑步、骑自行车等之间的关系。这些研究表明,“5D”变量是分析城市设计与积极的运动方式之间关系的关键。“5D”变量包括密度(density)、多样性(diversity)、步行空间设计(the Pedestrian Space Design)、目的地可达性(destination accessibility)、公共交通站点距离(distance to transit)。选择步行和乘坐公共交通(与更高程度的体育活动有关的因素)的人数,会随着密度、多样性、设计、可达性、换乘的可达性的提高而增加。这五个要素对于通过城市规划策略促进积极的运动有着至关重要的作用。

2.2 老年人的特征与健康需求分析

老年人的身体各项机能都出现退化。人体的各种器官都会衰老,如感官机能听觉、嗅觉、视觉触觉等会退化,认知能力、记忆力等会减退,肢体的灵活度、肌肉力量会下降。这些生理机能的衰退会导致老年人的公共休闲活动具有特殊性。

超重或肥胖的潜在原因,是缺乏运动和不健康的饮食,肥胖增加了患心血管疾病的风险。肥胖也和患一些常见的、致命的癌症的风险有关,比如乳腺癌和结肠癌。适量的体育活动除了有助于降低体内多余的脂肪,还能缓解紧张、焦虑、孤独的情绪。

2.3 基于老年人健康需求的新“5D”变量构建

① 密度(Density)

在对澳大利亚洛克汉普顿市 1200 名老年人的研究中,发现距离最近的公园超过 500m 的老年人日均活动量达到建议水平的,要比距离最近公园小于 500m 的老年人高出 40%。此外,相对于那些无法直接到达最近街头公园的老年人来说,可直接到达最近街头公园的老年人日均活动量要低 40%。

② 多样性(Diversity)

多样性反映了社区满足老年人的娱乐、文化、购物等多种活动需求的程度,以及老年人公共空间功能和设施的复合程度。

③ 步行空间设计(The Pedestrian Space Design)

步行空间设计反映了道路系统对老年人特殊的身体机能和心理特征的适应度。识别度、尺度感、多样性、安全性、围合感是评判步行空间设计好坏的五个要素。

④ 可达性(Destination Accessibility)

可达性指老年人从住宅到小游园的难易程度,以及设施与老年人的身材大小、姿态、操作能力的契合程度。可达性是老年人进入公共空间进行活动的前提,道路应便捷易识别,公共空间中设施的位置、布局、材质需迎合老年人的生理及心理特征。

⑤ 容错性(damage reduction)

由于老年人日常活动的范围集中在社区中,外出乘坐公交的几率相对较少,因此公共交通站点距离(Distance To Transit)在本文研究中不予研究。对老年人来说,除上述四个变量外,容错性(Damage Reduction)是社区公共空间不可或缺的要素,容错性反映了公共空间设计降低偶发事故或无意识行为对老年人造成伤害的程度。

2.4 本章小结

公共健康与城市设计的关系一直被学者们研究与探讨,在纽约政府颁布的纲领性文件:积极设计导则中,首次提出了分析公共健康与城市设计关系的关键——“5D”变量,分别从五个方面对纽约的公共健康问题提出解决策略。对于中国老龄化社会的基本国情,首先要了解老年人的特征,包括生理特征、心理特征、行为特征及健康需求。基于老年人的特征及健康需求,在“5D”变量基础上提出适用于中国老龄化社会的新“5D”变量,具体包括:密度、多样性、步行空间设计、可达性、容错性,为通过城市设计手法解决中国老年人健康问题提供一些参考和依据。

3 北京社区公共空间适老性案例研究

3.1 调研的内容、方法和对象的选取

本文对北京三个较典型的社区的老年人户外活动进行调研。考虑到小游园原本就匮乏的老旧社

区,以及一些无法满足居民基本活动需求的新建社区,很难反映社区的公共空间已建成环境与老年人健康需求之间的关系。因此本文选取了三个 2000 年以后建成的具有一定规模的中、高档社区进行调研。

影响老年人在社区公共空间进行活动的主要因素有性别、年龄、经济收入、文化水平等,本文通过实地调研、问卷调查、现场访谈的方式进行研究。另外,由于气候条件对老年人户外活动的的时间和频率影响较大,所以分别在春、夏、秋、冬四个季节对老年人进行调研。

3.2 调研结果研究分析

对于高峰季节各类户外活动的人数分析可看出,老年人更偏爱如聊天、静坐、带小孩等静态活动,其人数甚至占到了所有活动的绝大部分比例。对于老年人休闲活动的时间与空间分布规律进行分析,老年人在上午 8~10 点、下午 3~4 点和晚上 20~21 点形成三次休闲活动的高峰,各时段对于公共空间的需求具有明显的差异。对步行空间设计的老年人需求分析可看出,不同收入水平的老年人出行方式略有差异,但具有共同特征:以步行为主,其次是公交车、自行车、地铁、出租车,使用摩托车及私家车的老年人极少,由此可见优化步行空间设计对老年人日常出行至关重要。社区的出入口空间可达性的好坏是影响老年人能否进入到户外公共空间的首要条件,由于老年人易疲劳的特征,良好的坐憩空间是老年人参与到公共空间活动中的前提。

3.3 社区公共空间普遍存在的问题分析

① 由于老年人基本以静态活动为主,老年人的健康主要受益于往返小游园与住宅间的步行,因此小游园过密集不利于老年人健康;

② 社区功能过于单一,无法满足老年人购物、娱乐、邮寄等需求;

③ 社区中的步行小径缺乏设计,或单调冗长,或尺度不合理,现状环境使老年人无法每天坚持进行步行锻炼;

④ 社区公共空间中道路系统识别度不高,公共空间常用家具材质、布局及尺度不合理,使公共空间的可达性不足;

⑤ 社区公共空间容错性有待加强,缺乏容错性的公共空间设计易给老年人带来伤害。

3.4 本章小结

通过对北京三个典型社区的公共空间适老性情况进行调研,选取的对象均为 2000 年以后建成的社区,因为此阶段的社区在对公共空间设计有一定的考量,能反映出公共空间现状是否满足老年人的健康需求。通过调查问卷、实地观察、访谈等形式对高峰季节老年人各类户外活动的人数、老年人对社区步行空间的要求、对公共空间可达性的需求及公共空间安全性的需求进行了调研,发现目前社区主要存在的问题,社区公共空间的适老性设计尚存在诸多不足,为满足老年人的健康需求,本文的研究即针对上述设计不足,以新“5D”变量为切入点,探讨社区公共空间适老性设计原则与策略。

4 基于新“5D”变量的社区公共空间适老性设计探究

4.1 新“5D”变量在社区公共空间适老性设计的应用

对于密度,本文基于多项研究成果认为过密的公共活动空间不利于降低老年人 BMI 指数,老年人在公共活动空间中基本以静态活动为主,小游园分布密集就意味着缩短了老年人从住宅到公共活动空间的步行距离,减少了利于老年人健康的动态活动。

提高社区功能多样性有利于促进老年人健康,社区拥有越多功能的设施意味着老年人步行的机会

增加,与此相反的是,一块小游园用地只承载一种功能比集合舞剑、棋牌、太极拳等多种功能更吸引老年人前往,小游园应功能单一但设施多样。

评判步行空间好坏的标准有五个,分别为:识别度、尺度感、多样性、安全性、围合感,对每项进行合理的设计能更好的优化步行空间。

公共空间的可达性是老年人进行户外活动的前提,可达性的计算与两个要素有关:交通成本及断点区位吸引力。步行空间应互相连通成环路使老年人便捷地进入相应的活动区,断点区位吸引力主要指公共活动空间的设施满足老年人生理、心理特征及健康需求的程度,主要从坐憩设施、健身设施、人工照明、楼梯间设计进行研究。

公共空间的容错性对老年人进行户外活动的安全至关重要,要注意场地环境及室外配套设施的容错性设计。

4.2 基于新“5D”变量设计原则的案例社区改进建议

根据新“5D”原则可知,并非距离公共绿地越近,老年人的 BMI 指数越低,因此在改进设计中将几处组团绿地的硬质铺装取消,合并为面积稍大的公共活动空间布置在社区中心绿地中,与社区主入口广场、社区中心绿地共同构成一条景观轴线。(图 1、2)



图 1 某社区景观改造 1

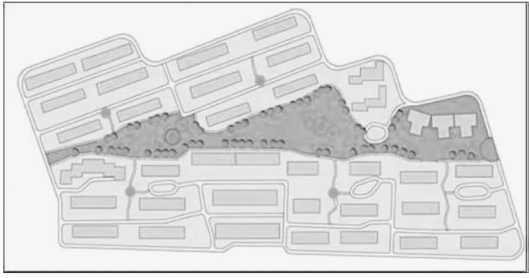


图 2 某社区景观改造 2

在改进设计中,将承载多样功能的公共建筑集中设在社区中心绿地中,可采用建筑底部为商业功能,上部为住宅住宅功能的商住混合形式。在社区东侧入口旁现为若干小餐馆,将其改造成小型农贸市场从而解决 A 社区及附近社区老年人购买新鲜食材难的问题。(图 3)

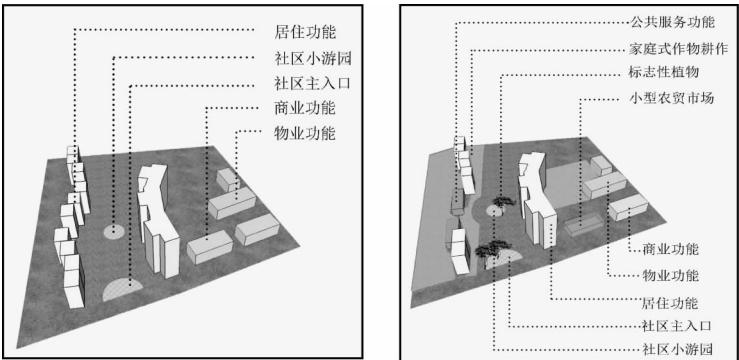


图 3 某社区公共绿地布局调整

老年人进行户外活动主要在入口广场、社区中心广场等有硬质铺装的场地中,但这些场地中的设

施过于单一,仅有几个座椅远远不能满足老年人的活动需求。以入口广场为例进行改造,在场地中增加花台式坐凳、休闲平台营造良好的休憩空间,增加景观小品、雕塑喷泉提供可供观赏的景观,在植物配置上不仅有规整的树阵,也有自由的景观树和密植式灌木,增加健身设施提供可供锻炼的场所。(图4)

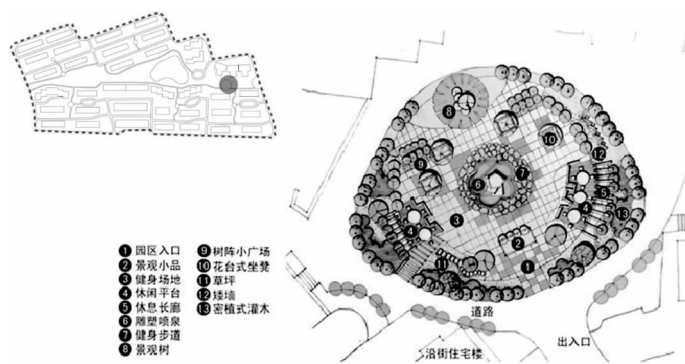


图4 某社区中心广场调整

通过在组团绿地中步行小径两侧增加设施创造个性鲜明、趣味性性强的步行空间。如沿着步行小径每隔100~125m设置折形休息座椅,供老年人步行途中休息;移去过于繁密的高大乔木,将这些遮挡视线的树木替换为树形、颜色各异的小乔木和灌木,不但形成丰富的视觉层次,而且创造了良好的围合感;在组团绿地节点设置景观休息亭或健身设施,丰富整条步行路径上的活动内容。(图5)

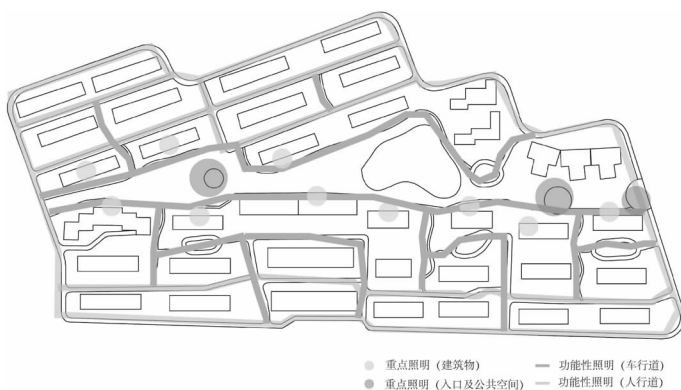


图5 某组团步行空间调整

在容错性方面,应在单元出入口增加坡道及扶手,方便行动不便的老年人出入。将老年人最经常前往进行活动的社区中心绿地中的硬质铺装改成软质铺装,采用橡胶或人造草坪等。社区缺少人工照明对老年人的夜间出行造成了一定的障碍,应对社区入口、社区中心绿地进行重点照明,对社区沿街建筑进行装饰照明,并对车行道、人行道分别进行功能性照明。

5 促进老年人健康的社区公共空间设计原则

安全舒适原则是社区适合老年人的公共空间设计的首要原则。随着年龄的增长,老年人的身体机能逐渐衰退,对空间的安全舒适性要求较高。空间设计应适应不同身高、体型、行动能力的老年人使

用,无论老年人坐着还是站着,都能有足够的体积和空间进行户外活动。

易达性原则指对于行动不便的老年人来说,社区的道路系统应简单、易识别,方便老年人到达。社区内的供老年人活动的公共空间指除了汽车,任何人都可以进入的区域,应设计一个可聚集的、可供休憩的、视线可达性好的公共活动空间来鼓励老年人进行户外活动。

可参与性原则指不同年龄段、不同身体状况的老年人都能公平地使用环境、设施、交通等。设计应考虑到老年人的心理特点,避免给老年人造成心理影响和危害。应重视老年人的想法和需求,鼓励老年人参与到设计中来。

地域性原则指自然环境、建筑风格、生活方式、审美情趣、传统民俗构成了独特的地域性文化,公共空间应迎合老年人的“怀旧”心理,能唤起老年人对城市的记忆和体验,从而帮助老年人减少传统文化与现实生活脱离带来的失落感和茫然感。

参 考 文 献

书籍:

- [1] 李德华主编.城市规划原理.北京:中国建筑工业出版社(第三版),2001
- [2] (丹麦)杨·盖尔著,何人可译.交往与空间[M].中国建筑工业出版社,1992
- [3] (日)芦原义信著,伊培桐译.外部空间设计[M].中国建筑工业出版社,1985
- [4] 胡仁禄、马光著.老年居住环境设计[M].东南大学出版社,1995
- [5] (西)Arian Mosteadi 编,杨小东,钟声译.老年人居住建筑[M].北京:机械工业出版社,2008
- [6] 周文麟著.城市无障碍环境设计[M].科学出版社,2000
- [7] 姚时章、王江萍著.城市居住外环境设计[M].重庆大学出版社,2000
- [8] 邓树勋著.老年人体育健身指南[M].华南理工大学出版社,1999
- [9] 田雪原著.中国人口学和中国人口老龄化[M].中国经济出版社,1991
- [10] (挪威)诺伯格·舒尔茨著,伊培桐译.存在·空间·建筑[M].中国建筑工业出版社,1990
- [11] (美)克莱尔·库珀·马库斯、卡罗琳·弗朗西斯著,俞孔坚、孙鹏、王志芳译.人性场所—城市开放空间设计导则[M].中国建筑工业出版社,2001
- [12] (英)D·肯特著,谢立新译.建筑心理学入门[M].中国建筑工业出版社,1992
- [13] 常怀生.建筑环境心理学[M].中国建筑工业出版社,1986
- [14] (日)野村欢著,北京市建筑设计院技术情报所摘.为残疾人及老年人的建筑安全设计[M].中国建筑工业出版社,1990
- [15] 邬苍萍著.社会老年学[M].中国人民大学出版社,1999
- [16] (美)戴安娜·Y·卡斯蒂斯著.针对老人的场地规划和设计:问题、导则和方案[M]
- [17] 开育岱.老年人建筑设计图说[M].山东科技技术出版社,2004
- [18] 李道增.环境行为学概论[M].清华大学出版社,1999
- [19] 刘连新,蒋宁山.无障碍设计概论[M].中国建筑工业出版社,2004
- [20] 王江萍.老年人居住外环境规划与设计[M].中国电力出版社,2009
- [21] 高保真,黄南翼.老龄社会住宅设计[M].中国建筑工业出版社,2006
- [22] (美)简·雅各布斯.美国大城市的死与生[M].译林出版社,2006
- [23] (日)高桥仪平著,陶新中译,牛清山译.无障碍建筑设计手册——为老年人和残疾人设计建筑[M].北京:中国建筑工业出版社,2003

- [24] Theo Beckers. Leisure and the Environment; Essays in Honour of Professor J A Patmore. Journal of Retailing and Consumer Services, 1995(3): 206
- [25] Chiuri, Maria Concetta. Individual Decisions and house hold demand for consumption and leisure. Research in Economics, 2000, 54(3): 227~324

期刊文章:

- [26] 李煜, 朱文一. 纽约城市公共健康空间设计导则及其对北京的启示[J]. 世界建筑, 2013, 9
- [27] 姚佳纯. 马斯洛需求理论下的现代社区规划设计思考[J]. 规划师, 2015, 229(31)
- [28] Nielsen, T. S. & Hansen, K. B. [2007]. Do green areas affect health? Results from a Danish survey on the use of green areas and health indicators. Health and Place, 13, 839—850
- [29] Ward Thompson & Aspinall, P. (2013). Woodland improvements in deprived urban communities: What impact do they have on people's activities and quality of life? Landscape and Urban Planning, 118, 79—89
- [28] 胡仁禄. 城市老年居住建筑环境研究概要[J]. 东南大学学报, 1994, 24, 6
- [29] 胡仁禄. 老龄化城市居住环境初探[J]. 建筑学报, 1994, 10
- [30] 张薇. 中国城市老年人居住问题的探讨[J]. 国外建材科技, 2002, 23(2): 84—86
- [31] 林文洁, 于喆, 杨绪波. 居住区老年人户外活动及其空间特征研究——夏季户外活动实态调查[J]. 建筑学报, 2011(20): 73—77
- [32] 郭玲. 无障碍设计在日本的实施. 世界建筑, 1984
- [33] 龚文晔, 白飏. 探讨适合老年人的居住区户外环境设计——以郑州 21 世纪社区为例[J]. 住宅科技, 2010(9): 33—35
- [34] 薛亮, 韩伟. 适应人口老龄化的居住区道路设计[J]. 山西建筑, 2007, 33(16): 28—29
- [35] 夏明, 武云霞. 人性化的老年人居住区设计[J]. 青岛理工大学学报, 2007, 28(5): 48—51
- [36] 詹林, 黄林青. 老年住宅建筑设计中的社会心理行为思考[J]. 重庆科技学院学报(自然科学版), 2006, 8(3): 44—46
- [37] 姚建平, 尹建强. 老年人居住区户外环境及景观设计刍议[J]. 湖南农业大学学报(自然科学版), 2010, 36(2): 38—40
- [38] 夏明, 杨丽武, 云霞. 老年人居住模式设计[J]. 城市问题, 2005(1): 46—49
- [39] 孙翠翠, 张瑞丽. 老年人对城市公共座椅和公共卫生间利用情况的调查研究[J]. 环境卫生学杂志, 2013, 3(3): 206—210
- [40] 吴义曲. 基于老年人心理需求的居住区景观设计[J]. 武汉理工大学学报, 2010, 32(15): 56—60
- [41] 韩立新, 柴寅. 房地产开发中的老年人公共空间人性化研究[J]. 山西建筑, 2008, 34(4): 6—7

学位论文:

- [42] 谷鲁奇, 面向老年人的旧住宅区公共活动空间更新方法研究[D], 重庆大学, 2010
- [42] 宋旭煌, 中国老年型城市老人社区生活空间环境研究[D], 同济大学硕士论文, 1996
- [43] 朱松, 城市老年人户外活动空间[D], 同济大学硕士论文, 1996
- [44] 王涛, 老年居住体系模式与设计探讨[D], 西安建筑科技大学, 2003
- [45] 尚冰, 老年公寓设计探讨[D], 大连理工大学, 2000
- [46] 赵鹏, 老年人居住环境空间需求研究[D], 西安建筑科技大学, 2001
- [47] 强虹, 适宜老年人的城市公共空间环境设计研究——以西安环城公园为例[D], 西安建筑科技大学, 2004
- [48] 梁娅娜, 居住区户外环境老年人适应性研究[D], 大连理工大学, 2006
- [49] 周海源, 西安市居住区老年人户外活动场所规划设计研究[D], 西安建筑科技大学, 2007
- [50] 孙莹, 城市居住区老年人休闲娱乐场所研究[D], 青岛理工大学, 2013

Research on appropriate public space based on public health in the context of population aging

Cao Yang

(School of Architecture and Art, North China Univ. of Tech., Beijing, 100144)

Abstract Most previous studies based on physiological and psychological characteristics of the elderly. This paper studies public space from a new viewpoint that the elderly public health. Research on the relationship between public health and urban design abroad mainly from relationship between landscape design and urban design. The active design guidelines which made by American in 2010 is a programmatic document. The guidelines proposes that five “D” variables are the key to analysis the relationship between public health and urban design. For the health needs of the elderly in China, this paper proposes new five “D” variables which are density, diversity, the pedestrian space design, destination accessibility and damage reduction. Explore the design principles and strategies, and propose suggestions for improvement of the studied communities.

Key words public health; public space; senior-fit; beijing

中国铜工业景气指数与预警系统研究

研究生 帅加琴

导师 吴振信 王中奎

(北方工业大学经济管理学院 北京 100144)

摘 要 铜在现代社会生产生活中扮演着重要角色,但目前对中国铜工业的监测预警却鲜有研究。本文运用合成指数、预警信号灯系统和粗糙集—神经网络模型对中国铜工业进行监测预警研究。首先,初选指标共 45 个,对指标进行预处理和考虑中国移动假日的季节调整后,经时差划分,最终确定 9 个先行指标、4 个一致指标和 4 个滞后指标。其次,编制了铜工业景气合成指数,实证表明,先行合成指数平均领先一致合成指数 2—3 个月,滞后合成指数平均滞后于一致合成指数 2 个月。然后,建立了预警信号灯系统,直观表示铜工业经济运行的冷热程度。最后,构建粗糙集——神经网络模型对铜工业的警度进行预测,预测 2015 年一月处于偏冷状态。

关键词 铜工业;景气指数;预警;粗糙集;人工神经网络

作者简介 帅加琴,女,1989 年出生,硕士研究生,主要研究方向为大宗商品定价及数量经济分析。

E-mail: shuaijiqin@126.com

1 绪论

1.1 研究背景

铜是具有战略意义的资源,在生产生活中应用广泛。铜工业作为基础原材料工业,其发展态势对国民经济的运行有重大影响。中国铜工业发展仍存在诸多问题,运行不稳定甚至出现剧烈波动^[1]。若建立铜工业的监测预警系统,可以预测预报行业景气情况,以采取防范措施。

1.2 相关文献综述

景气指数亦称景气度,是用定量的方法全面衡量经济波动的幅度^[2](高铁梅,2003)。经济预警建立在景气分析的基础上,是对警情的预测预报。经济景气分析可追溯到 19 世纪末,常用的景气指数有三类^[3](扩散、合成及 S—W 型景气指数),其中合成指数有其独特的优势,是国际上通用的效果最好的景气指数。传统信号灯方法简单直观^[4],但在预测及智能化方面存在一定缺陷。国内外学者对基于模式识别的经济预警理论与方法的研究取得了一系列成果^[5],对本文研究有很大的借鉴意义。

1.3 论文结构安排及创新点

论文主体包括景气指标体系构建、景气合成指数编制和预警系统开发三大部分:首先,论述本文研究背景,并对国内外相关理论及实证研究进行系统回顾。其次,建立景气指标体系。在确定指标体系的基础上,将景气指数、预警信号灯系统、粗糙集——神经网络模型结合起来对中国铜工业进行监测预警。最后是结论与展望。

主要创新点:首次对铜工业进行监测预警,研究成果对类似单品种金属工业的研究具有重要参考价值;其次,在数据预处理、基准指标以及增长率循环指数设计等方面都有所创新,使得指数的设计更符合中国铜工业现状。

2 中国铜工业经济监测指标体系的建立

指标体系是对行业进行经济监测预警的基础。

2.1 指标筛选

从大量经济指标中选择中国铜工业的景气指标必须遵循一些基本原则。本文指标体系的设计原则包括:经济上的重要性与互补性、统计上的充分性与及时性以及指标变动的灵敏与可靠性等。初选指标从以下逻辑来考虑:首先,宏观投资环境、政策等对铜工业的投资、生产等都有一定影响;其次,铜工业的下游应用市场主要以电力电缆、汽车零部件、家用电器及建筑等领域为主,因此这些产业的发展状况决定其对铜产品需求,这将间接影响铜的投资和生产。

根据铜工业监测预警系统指标体系的设计原则和对初选范围的确定,同时借鉴钢铁、煤炭等相关行业的监测预警经验,本文收集了 45 个指标,主体包括四大部分,即影响铜工业的宏观经济指标、生产资料供给、下游产业指标以及铜工业自身核心指标。

2.2 数据采集与预处理

选取 2004 年 1 月至 2014 年 12 月的月度指标,数据主要来源于中国有色金属工业协会、北京安泰科信息开发有限公司及国家统计局和新浪数据库。由于有色金属行业统计的历史缘由,指标数据存在大量缺失。本文对缺失值的处理结合了统计方法与行业内人士建议,主要应用了均值法、比值法和 EM 方法。

2.3 考虑中国移动假日的季节调整

铜工业的监测预警采用的是月度数据,而且需要环比增长率序列,所以必须对经过前期预处理后的月度数据进行季节调整。本文对指标数据的季节调整考虑了移动假日的影响,主要是借鉴三区段变权重春节模型,通过 EViews 编程实现。季节调整可滤去季节因素,考虑移动假日后,在每年的 1 至 3 月份能显示出细微差异。

3 中国铜工业景气合成指数的研制

合成指数兼顾各指标的波动状态和波动程度,既可反映经济波动幅度又可预测转折点。鉴于此,本文采用合成指数方法来编制中国铜工业的景气指数。

3.1 合成指数原理简介

单个经济变量的波动无法代表行业整体的波动,为了正确地反映和预测铜工业经济波动状况,必须综合考虑各相关领域的景气变动及相互影响,制定综合指数。合成指数是经济监测预警的核心方法之一,其本质是多指标的加权平均。

3.2 基准指标及基准循环的确定

经济指标本身所代表的经济活动错综复杂,可能先行、一致或滞后于行业经济周期波动。划分指标组,首先需要确定一个基准指标。学者们在进行景气研究时一般采用研究对象的总产值作为基准,但铜工业作为一个细分的行业尚缺这个指标。经过比较研究及业内人士建议,我们根据当月(精铜产量 \times 当月均价)计算精铜产值,将其作为基准指标,并用增长率循环。

3.3 先行、一致、滞后指标的划分

时差划分要求先行关系稳定,其指标峰值比基准峰值先行 3 个月以上;滞后指标选取标准和先行

指标相同;一致指标与基准保持在 2 个月内的时差。经过时差相关分析与峰谷比对,结合铜工业的实际情况及指标的经济含义,最终确定出先行、一致和滞后指标组,具体见表 3-1。

表 3-1 中国铜工业景气指标划分组(“-”表示先行,“+”表示滞后)

类型	指标名称	先行滞后期数	时差相关系数
先行指标(9)	LME 铜价	-4	0.6363
	M2	-4	0.6083
	PPI	-3	0.5704
	有色金属采矿业固定资产投资	-5	0.7067
	房地产销售面积	-4	0.5634
	发电机组产量	-5	0.5960
	集成电路产量	-3	0.6303
	家电组合产量	-3	0.5783
	汽车产量	-4	0.4902
一致指标(4)	精铜产值	0	1
	铜矿进口量	-2	0.4116
	主营业务收入	+1	0.7992
	主营业务成本	+1	0.7316
滞后指标(4)	产成品占用额	+3	0.6408
	流动资产平均余额	+4	0.5704
	应收账款总额	+4	0.6645
	精铜出口量	+4	0.5547

3.4 合成指数的计算及分析

先行、一致和滞后各指标本身的波动能够部分反映该领域情况,但要观察行业经济波动的整体,必须综合各个变量的波动。合成指数的计算是先求出各指标的对称变化率,再计算先行、一致和滞后三组指标的组内、组间平均变化率。最后,以某年为基年,计算出其余年份每月的相对指数。本文参考的是 NBER 的标准算法,其中指标权重的设定用了熵值法,并且采用 HP 滤波法进行去趋势调整。

按上述方法计算出中国铜工业的先行、一致和滞后合成指数,结合图形显示对经济运行状况进行分析,其对应曲线为图 3-1。

从图 3-1 可以看出,先行、一致、滞后合成指数三者的关系是比较明显的:先行合成指数的峰谷平均领先于一致合成指数 2-3 个月,滞后合成指数的峰谷平均滞后于一致合成指数 2 个月。

(1)铜工业发展的景气分析:中国铜工业一致合成指数从 2004 年初低位上升后,至 2008 年初一直保持在正常区间内震荡。自 2008 年二季度开始大幅下降,并于年底降至数年来最低谷;随后再次出现快速回升态势,并保持在较高位波动。2014 年中期至年末呈下降趋势。

(2)铜工业发展动向分析:从图 3-1 可以看出,先行合成指数在 2014 年 5 月份开始下降,且没有回升,这预示着可能还会继续下降。参考其平均超前期数,可以认为铜工业在 2015 年中期增速仍将保持下降趋势。中长期来看,随着中央稳增长和调结构相关政策的贯彻实施,国内原材料需求总体有望稳定增长,铜工业亦有望企稳发展。

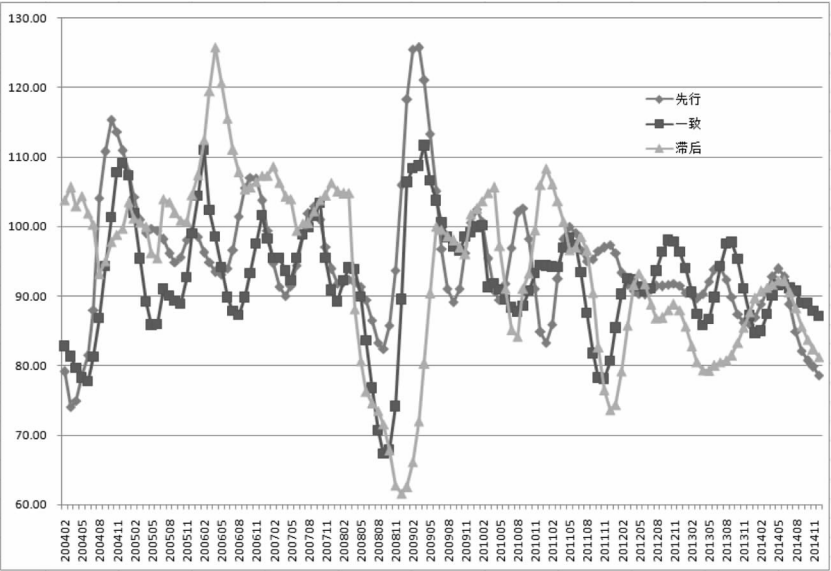


图 3-1 先行、一致、滞后合成指数图

4 预警信号灯系统的建立

4.1 预警指标及权重

鉴于在编制景气指数时已对指标进行了严格筛选和预处理,因此将先行指标和一致指标都考虑进来作为预警指标组。为计算综合预警指数,再次结合熵值法和专家意见对各指标赋权。

4.2 预警界限的确定

预警分析中将经济运行状态分为五种,因此,必须有四个临界值。本文根据 3σ 法和中经景气监测中心的思路,采用有色协会专家商定的概率值(红至蓝灯区分别占 8%、14%、56%、14%、8%),确定临界值。划分指标所处区间,并赋予相应分值。红、黄、绿、浅蓝、蓝灯区各赋 5、4、3、2、1 分,最后加权得到综合分数。标准化后得到的序列即是综合预警指数,并给出对应灯号。

4.3 预警信号的输出及分析

根据指标临界值的分布,客观地划分出中国铜工业的预警状态。2014 年初铜工业景气综合状况出现过冷(1、2 月)和偏冷状态(3 月),年中大部分时间保持在正常区间(6—10 月),但是年底出现偏冷信号(11、12 月)。回顾合成指数图,可以看出预警信号灯与指数图反映的趋势基本吻合。

5 基于粗糙集——神经网络的中国铜工业预警系统研究

景气指数和预警信号灯系统对中国铜工业的监测预警已经能够取得良好的效果,但都是基于线性结构模型,在解决经济系统复杂性、智能化方面较为困难。为了提升预警模型的性能,本章引入粗糙集理论和神经网络方法。

5.1 粗糙集——神经网络预警方法基本原理

人工神经网络模型可用来预测经济走势,但易陷入局部最优等问题。而粗糙集能够弥补这些不足,二者优势互补。

5.2 粗糙集——神经网络模型对中国铜工业预警的实证分析

将预警灯系统中的 13 个指标作为条件属性、预警指数作为决策属性,并利用粗糙集应用软件 ROSETTA 对数据进行离散化处理和知识约简。最终约简后的指标为 LME 铜价、M2、有色采矿业投资、集成电路产量、精铜产值、铜矿进口。从经济意义来看,属性约简后的指标具有代表性。

再将 131 个样本分为训练集、检验集和预测集,各 112 组、13 组和 6 组。正如神经网络原理所述,第 n 期的输入值对应的期望目标输出值则是第 $n+1$ 期的预警状况,这样网络训练得到的输出值具有预测下一期预警状态的功效。

(1)神经网络模型的确立:尚需确定的是隐层节点数。参考经验公式并反复试验,最终确定隐层节点数为 12,则网络结构为 6—12—5。

(2)模型的训练及验证:将 2004 年 2 月至 2013 年 5 月的 6 个指标作为输入,接下来一个月的警度作为输出值,得到最优化的网络权值矩阵。结果表明,在检验集的 13 组中,RS—NN 模型预测错误的只有一组。为了说明本文 RS—NN 模型的性能,同时仅用 BP 模型进行预警,精度分别为 92.86% 和 78.57%。

目前还没有哪一种预警模型是完全最优或通用的,可能都有一定的局限性。相对而言,本文结合 RS 和 BP 的预警模型具有可行性和实用性。

(3)预警结果分析:模型通过检验之后,将 2014 年 12 月的指标数据输入模型,对 2015 年 1 月份进行预警,其预警输出值表明该月处于偏冷状态。

6 结论与展望

本文主要在以下方面进行了研究:(1)构建了中国铜工业景气指标体系。初选指标共 45 个,经进一步时差划分,得到先行指标 9 个、一致指标 4 个、滞后指标 4 个;(2)编制了中国铜工业合成指数,实证表明:一致合成指数平均滞后于先行合成指数 2—3 个月,而平均领先于滞后合成指数 2 个月;(3)建立了预警信号灯系统。本文所建立的预警灯系统能够较准确地判断铜工业经济的冷热状况。2014 年初铜工业景气综合状况出现过冷和偏冷状态,年中大部分时间保持在正常区间,年底出现偏冷信号;(4)构建 RS—NN 模型对未来警度的预测。将粗糙集和神经网络优势互补,丰富了智能预测技术在经济监测预警领域的应用。本文建立的 RS—NN 模型对中国铜工业警度有着良好的短期预测能力,且预测 2015 年 1 月将处于偏冷状态。

本文在铜工业监测预警方面取得了一定的研究成果,但在指标体系设计、预警界限的确定方及智能模型预测部分都有待更广泛、深入地探讨。

参 考 文 献

- [1] 张方方等. 试析中国铜工业存在的问题与对策建议[J]. 中国矿业, 2013, 22(2): 9—13
- [2] 高铁梅, 孔宪丽, 刘玉. 中国钢铁工业景气指数的开发与应用研究[J]. 中国工业经济, 2003(11): 71—77
- [3] 王金明等. SW 型先行景气指数建设的实证研究[J]. 中国管理科学, 2007, 15 (4): 116 — 123
- [4] 王德青, 孙玲玲, 王翔. 中国有色金属工业经济预警系统研究[J]. 中国矿业, 2010, 19(9): 25—29
- [5] Mukta Paliwal. Neural networks and statistical techniques: A review of applications[J]. Expert Systems with Ap-

The Study of China's Copper Industry Composite Index and Early Warning System

Shuai Jiaqin

(School of Economics and Management, North China Univ. of Tech., Beijing, 100144)

Abstract Copper plays an important role in the modern society, and its development trend will have a great impact on the national economy. In this paper, we use composite index, early warning system and RS-NN model to monitor the copper industry. Firstly, we choose a total of 45 primary indicators, and finalized nine leading indicators, four consistent indicators and four lagging indicators. Secondly, we establish the composite index and the research shows that the turning points of consistent composite index have averaged a 2–3 months lag over the leading one, two months lead over the lagging index. Then, establish the early warning system. Through empirical research on China's copper industry from Jan. 2004 to Dec. 2014, we conclude that the established warning system can accurately estimate the industry conditions. This paper also constructs the RS-NN model to predict the warning degree of the copper industry, in which, artificial neural network model for predicting the future trends, and the rough set is used to simplify the input layer of artificial neural network model to enhance early warning accuracy. The RS-NN model can be used in short-term prediction or a longer period, and determine the warning degree.

Key words copper industry; composite index; early warning; rough sets; artificial neural network

我国铝业生产能量流研究

研究生 刘实实

导师 赵贺春

(北方工业大学经济管理学院 北京 100144)

摘 要 铝业生产过程是高能耗的,铝行业的高速发展态势给我国资源、能源和环境带来了巨大冲击,因此挖掘铝业生产的节能潜力成为现阶段我国铝工业亟待解决的问题。本文通过深入的数据收集整理工作,首先细致地描述了我国铝业生产能耗现状,对我国铝业生产能源消耗水平、能耗结构等方面进行详细阐述。然后运用扩展的三环节结构模型构建出我国铝业生产实际能量流静态图景,并对具体能耗指标进行精确测算。最后运用灰色 GM(1,1)模型对我国未来十年吨铝综合能耗变化趋势进行预测,结果表明基于 GM(1,1)模型构建的预测模型预测精度较好,其真实值与预测值拟合度较优,预测结果高度可靠,能够反映我国吨铝综合能耗的客观存在和发展态势,能够为我国制定宏观能源决策提供可靠依据。

关键词 铝业; 能量流; 综合能耗; 灰色 GM(1,1)模型

作者简介 刘实实,女,1990 年出生,硕士研究生,主要研究方向为环境会计。

E-mail: liushishi_ok@163.com

1 绪论

铝的生产一般需要经过采矿、氧化铝生产、原生铝电解以及原生铝液铸锭等主要工序,每一道生产工序都需要能源的分配和转换。作为“永不消逝的金属”,铝的生产过程能耗很高,每吨铝平均耗电量约为 1.2—1.3 万千瓦时/吨,在电解铝环节,平均综合耗电量为 14000 千瓦时/吨(新标准),同时,氧化铝在生产过程中每吨耗能 0.5 吨标准煤(间接加热拜耳法),能耗也是很高的。因此在节约能源、减少排放、保护生态、防止全球气候变化的过程中,铝是首选金属材料已成为共识。同时,铝行业的高速发展态势给我国资源、能源和环境带来了巨大冲击,因此挖掘铝业生产的节能潜力,控制铝业生产的污染物排放成为现阶段我国铝工业提高能源的利用效率、降低环境负荷亟待解决的问题。

2 我国铝业生产能耗现状分析

2.1 我国铝业生产能源消耗水平

我国铝锭综合交流电耗从 1980 年 17146kWh/t,下降为 2014 年 14323kWh/t,氧化铝综合能耗也从 2002 年每吨氧化铝能耗 1154.86kg(标煤)下降至 2014 年每吨能耗 617.03kg(标煤)。我国铝工业在构建生态铝业、绿色铝业的发展道路中取得了巨大成就。但在生产技术、资源利用效率方面,我国较国际先进水平仍有一定差距。第一,原铝生产技术比国外相对落后。电解能耗高于先进水平约 600kWh/t,电流效率低于先进水平 2%左右;第二,氧化铝单位能耗较国外单位能耗高出 1—3 倍。由于我国铝土矿 A/S 相对较低(一水铝石),矿石溶出难度大,溶出温度需达到 260℃左右,因此氧化铝企业更多的采用烧结法、混联法生产,工艺复杂,能源消耗高,造成氧化铝生产单位能耗较高。而国外大

多数矿石为三水铝石,溶出性能好,溶出温度仅需 150℃左右,因而国外氧化铝企业更多的采用拜耳法生产,相对氧化铝单耗较低。

2.2 我国铝业生产能耗结构状况

统计数据显示,1995 年到 2013 年,我国铝业生产能源消耗总量从 2379.23 万吨标准煤增加至 11823.47 万吨标准煤,增长了 496.89%,如下图 2-1 所示。

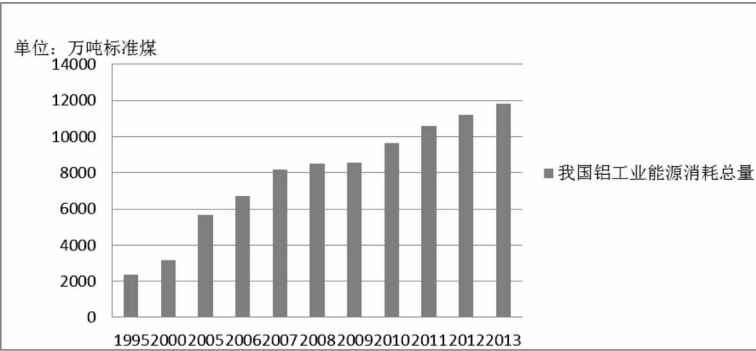


图 2-1 铝工业能源消费总量趋势图

从消费结构角度分析,我国铝业生产能源以煤炭和电力为主。2013 年,在铝业生产中,煤炭和电力的能源消费量占总消费量的 70%以上,其中煤炭占 42.15%、电力占 30.58%。下面将展开分析我国铝业生产的能源消费结构。

(1)煤炭能源消费

1995 年至今,铝业生产中煤炭消费量在起伏中呈总体上升趋势,数据表现为,1995 年煤炭消费量为 1066.31 万吨标准煤,2013 年煤炭消费量增至 4718.92 万吨标准煤,平均年增长 13.78%(见图 2-2)。然而,2009 年的铝工业煤炭消费量呈现负增长态势。根据统计数据,2009 年我国铝业生产煤炭消费量为 2226.63 万吨,较上年下降 6.39%。

从煤炭消费比重角度分析,1995 年至 2013 年我国铝业生产煤炭消费比重呈相对下降趋势。2013 年,铝业生产煤炭消费总量占铝工业用能总量的 26.25%,比 1995 年下降 21.64%,比 2009 年下降约 0.47%。综合分析,由于水电的发展,能源价格变化,以及铝工业生产技术进步,1995 年至 2013 年,我国铝业生产煤炭消费总量呈上升趋势,而煤炭消费比重则呈下降趋势。

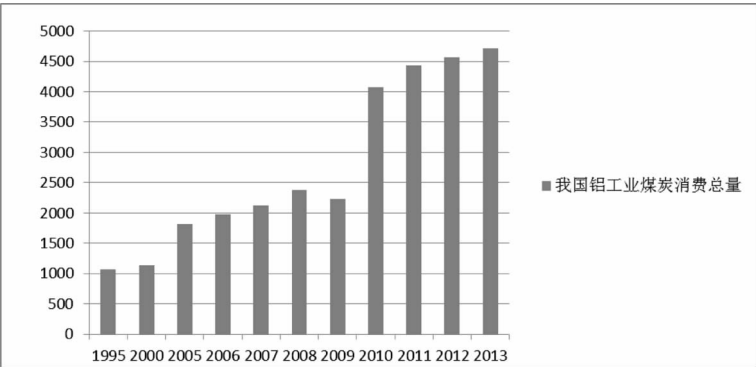


图 2-2 煤炭能源消费趋势图

(2) 电力能源消费

铝具有很高的反应性,与氧具有很强的化学结合力,因此需要大量能量才能将氧化铝(Al_2O_3)中的氧和铝元素分离。一般情况下,生产一吨原铝需要 130—150 万千瓦时交流电。在铝工业生产过程中,各个环节都要消耗大量的能源,但电力能源的消费主要集中在电解还原过程中,电解铝电力能源消耗约占整个铝业生产能耗的 80%(见图 2—3)。

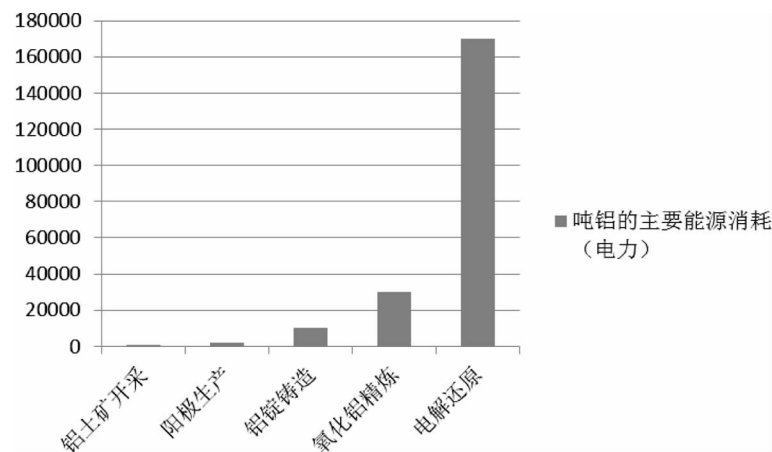


图 2—3 吨铝主要能源消耗图

目前,国际铝工业用电中普遍使用相对廉价的水电,水电占全球铝工业用电结构的 67%。具体来看,在加拿大和独联体,原铝生产 100%都使用水电;美国铝工业用电结构中,水电比重高达 70%;而在委内瑞拉、巴西、挪威,铝工业使用水电的比重也是非常高的。作为借鉴,我国在发展铝工业的过程中,可以因地制宜地利用不同地区电力资源的差别,采取多种电力能源。例如,在自然经济条件较差、但拥有丰富水力资源的边远地区,运用“资源转换”战术调整我国铝工业电力结构。我国铝业生产电力结构如图 2—4 所示。

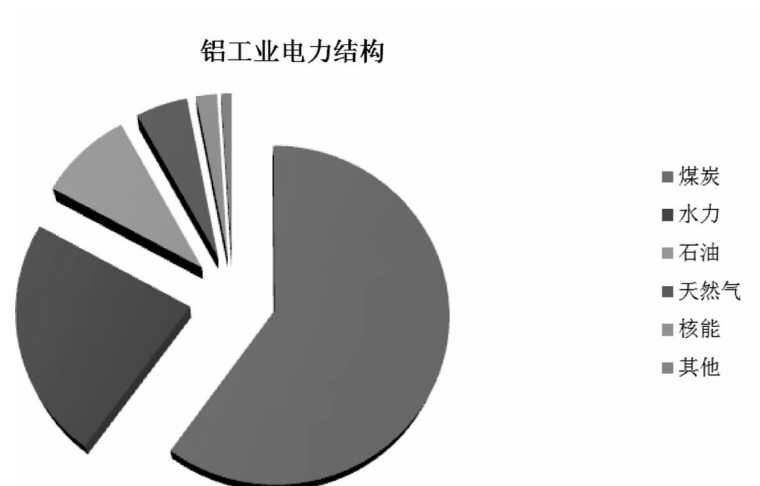


图 2—4 铝业生产电力结构图

2013 年铝业生产中铝锭综合交流电耗下降到 13479.31 千瓦时/吨,比上年同期下降 188 千瓦时/吨,节电约 31 亿千瓦时,占铝业生产用能 30.58%,已到达世界先进水平。总的来讲,2013 年铝锭综合交流电耗比 2005 年下降 1079 千瓦时/吨,节电约 174 亿千瓦时,减少二氧化碳排放 721 万吨。1995 年以来,铝业生产用电量总体呈现上升趋势,平均年增长为 4.17%。不过,2005 年我国铝业生产耗电量与之前相比明显出现增速放缓现象(见图 2—5)。

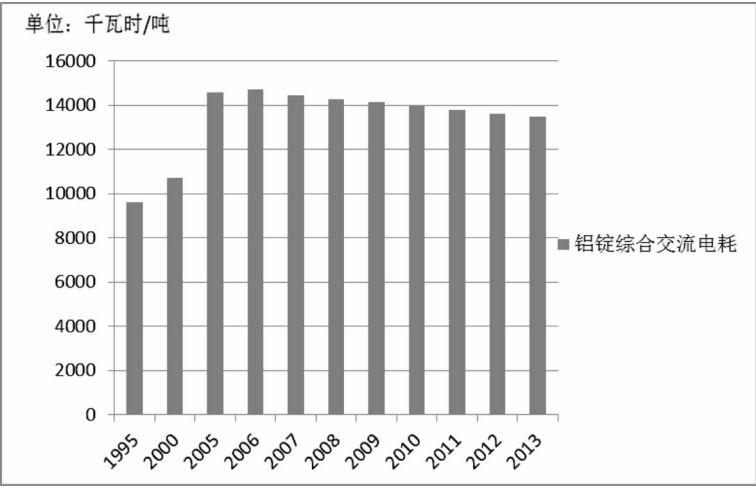


图 2—5 铝锭综合交流电耗趋势图

3 我国铝业生产能量流静态分析

3.1 相关数据收集和准备

本文主要分析在我国应用范围较广、基于拜耳—烧结联合法下铝工业的生产能量流。联合法一般可以分为并联、串联和混联三种工艺流程,本文着重分析并联法。

生产原料主要包括:煤、电、水、石灰、铝土矿和碳酸钠。

各环节主要产品包括:氧化铝、液铝等。

主要的副产品包括:一水硫酸钠、蒸汽、铝酸钠溶液、炭素材料。

如表 3—1 所示,本文以吨铝为统计基础,相关能耗指标均来自《中国有色金属工业年鉴 2014》中的主要经济技术指标、《能源统计年鉴》中的相关能耗数据、国家统计局及有色金属工业协会铝业分会的官方网站披露。文中涉及的一些非统计性工艺数据,则来源于相关研究文献和对典型铝工业企业主要技术指标的调研。

3.2 我国铝业生产实际能量流静态图景

根据我国铝业生产特点,结合相关能耗指标,可以大致描绘出我国铝业生产能量流静态图景。以扩展的三环节模型为基础,将铝业生产这一大系统划分为使用、转换、分配和回收四个子系统。

表 3—1 能耗数据

项目名称	具体指标	数值
产成品(10 ⁴ t)	氧化铝	2906.5
	原铝(电解铝)	1624.4
相关能耗	矿耗(kg/t)	2120
	石灰(kg/t)	260
	原煤(kg/t)	374.38
	碱(kg/t)	83.74
	絮凝剂(kg/t)	0.465
	电耗(kWh/t)	282.9
	蒸汽(kg/t)	1060
	水(kg/t)	1565
其他	原铝氧化铝单耗(kg/t)	1929
	原铝炭阳极单耗(kg/t)	498.63
	原铝氟化盐单耗(kg/t)	21.9
	氧化铝总回收率(%)	75.4

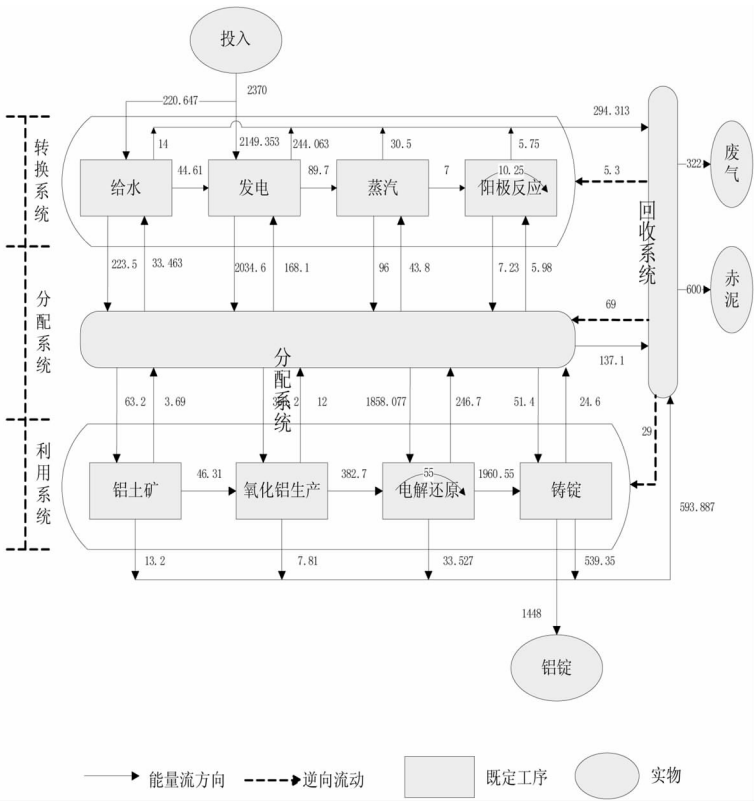


图 3—1 我国铝业生产能量流图

四个系统的具体能量流动状况如图 3—2 所示。在铝业生产企业能量流图中,企业投入的各种能

源经过转换系统、分配系统、使用系统、回收系统最后废物排放,如图 3—1,按照上中下的顺序 3 个实线框分别是能源的转换环节、能源分配环节和利用环节。各实线框内,每个箭头上的数据表示其代表的能量流大小(单位是 kgce),以 1 t 铝为基准。如图所示,生产 1 t 铝,投入到铝业生产企业的能源为 2370 kgce,能源损失为 922 kgce,生产过程中的实际消耗为 1448 kgce。

3.3 我国铝业生产能耗指标测算分析

铝业生产企业能耗指标测算包括计算吨铝综合能耗、能源效率和能源环境效率。计算能耗指标之前,首先要总结描绘铝业生产实际能量流框架,根据图 3—1,可以测算得出铝业生产企业各个能量流系统的输入输出情况,如图 3—2 所示。

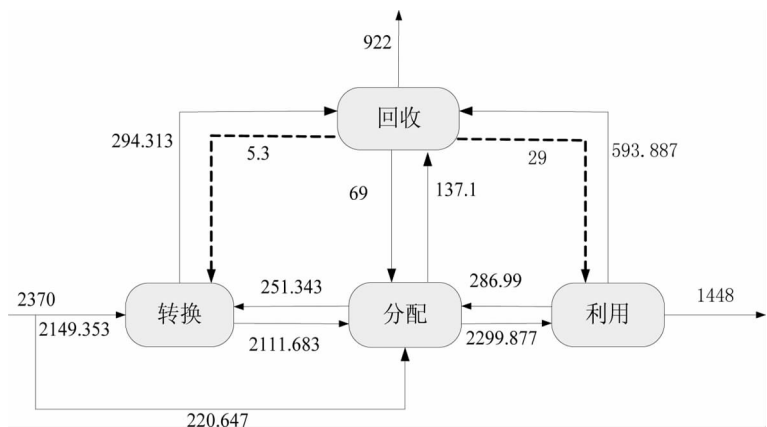


图 3—2 实际能量流框架

铝业生产企业的能量流分析框架中的相关参数及数据收集如表 3—2 所示。

表 3—2 相关参数及数据收集

能耗指标	数值
E_0	2370
λ	9.31%
λE_0	220.647
$(1-\lambda)E_0$	2149.353
E_1	2111.683
E'_1	251.343
E_2	2299.877
E'_2	286.99
E_{r3}	294.313
E'_{r3}	5.3
E_{r4}	593.887
E'_{r4}	29
E_{r5}	137.1
E'_{r5}	69

能耗指标	数值
E_{r6}	922
E_m	1448

在图 3-2 中,各个子系统(转换、分配、利用、回收)相互关联、影响和制约,并且每个系统都满足物质守恒定律和能量平衡定律。各个子系统的能量平衡情况如下:

在转换系统中,输入能量流包括 $(1-\lambda)E_0, E'_1, E'_{r3}$ 三股能量流,能量输入 $= (1-\lambda)E_0 + E'_1 + E'_{r3} = 2149.353 + 251.343 + 5.3 = 2405.996$; 输出能量流包括 E_1, E_{r3} 两股能量流,能量输出 $= E_1 + E_{r3} = 2111.683 + 294.313 = 2405.996$ 。满足转换系统能量平衡等式 $(1-\lambda)E_0 + E'_1 + E'_{r3} = E_1 + E_{r3}$ 。

在分配系统中,输入能量流包括 $E_1, \lambda E_0, E'_2, E'_{r5}$ 四股能量流,能量输入 $= E_1 + \lambda E_0 + E'_2 + E'_{r5} = 2111.683 + 220.647 + 286.99 + 69 = 2688.32$; 输出能量流包括 E'_1, E_2, E_{r5} , 能量输出 $= E'_1 + E_2 + E_{r5} = 251.343 + 2299.877 + 137.1 = 2688.32$ 。满足分配系统能量平衡等式 $E_1, E_2, E_{r5} = E'_1 + E_2 + E_{r5}$ 。

在利用系统中,能量输入 $= E_2 + E'_{r4} = 2299.877 + 29 = 2328.877$; 能量输出 $= E'_2 + E_{r4} + E_n = 286.99 + 593.887 + 1448 = 2328.877$ 。满足利用系统能量平衡等式 $E_2 + E'_{r4} = E'_2 + E_{r4} + E_n$ 。

在回收系统中,能量输入 $= E_{r3} + E_{r4} + E_{r5} = 294.313 + 593.887 + 137.1 = 1025.3$; 能量输出 $= E'_{r3} + E'_{r4} + E'_{r5} + E_{r6} = 5.3 + 29 + 69 + 922 = 1025.3$ 。满足回收系统能量平衡等式 $E_{r3} + E_{r4} + E_{r5} = E'_{r3} + E'_{r4} + E'_{r5} + E_{r6}$ 。

同样的,从整体角度出发,根据能量守恒定律、铝业生产实际能量流动分析框架(图 3-2)及各系统平衡关系,总的能量投入也等于总能量输出,即满足等式 $E_0 = E_n + E_{r6} = 1448 + 922 = 2370$ $E_0 = E_n + E_{r6} = 1448 + 922 = 2370$ 。

以实际能量流图(图 3-1)为基础,可以绘制出铝业生产基准能量流图(图 3-3)。图 3-3 中,铝业生产基准能量流图是一张“全封闭单行道”式的生产流程能量流图。其中满足以下两个假设:

- (1)在所有生产流程中,能量流流向均为从上游工序向下游工序流动;
- (2)在一次能源流转中,不涉及系统外的能量流输入与输出。

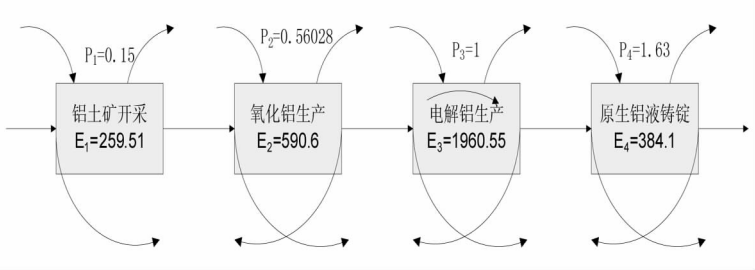


图 3-3 铝业生产基准能量流图

图 3-3 中, $P_1 - P_4$ 分别表示原铝生产过程铝土矿开采、氧化铝生产、电解铝生产以及原生铝液铸锭等主要生产工序的铝比系数,即第 i 道工序生产的合格产品量(液体产品以 m^3 计算,固体产品以 t 计算)占最终合格铝产量的百分比,单位为工序产品量/ $t-Al$ 。 $E_1 - E_4$ 表示统计期内各工序(第 i 道工序)的工序能耗。按照 e-p 分析法,吨铝综合能耗等于统计期内的铝工业企业所耗能源总量除以同期产量。

根据铝业生产基准能量流图,可以计算出吨铝综合能耗:

$$E_p = \sum_{i=1}^n e_i p_i = E_1 \times P_1 + E_2 \times P_2 + E_3 \times P_3 + E_4 \times P_4 = 259.51 \times 0.15 + 590.6 \times 0.56028 + 1960.55 \times 1 + 384.1 \times 1.63 = 2956.46$$

吨铝综合能耗 2956.46Kgce/t 折合约 21057.4kwh/t,这与《有色金属统计年鉴 2014》中的数据相吻合。

本文收集到的企业能量流相关数据如表 3-2 所示,这些数据初步表明了企业能量流的大体状况,同时将进一步分析铝业生产企业的主要能耗指标(表 3-3)。分析能耗指标的作用主要是通过评价企业整体能耗现状,与现有先进标准比较从而找出差距,挖掘节能潜力。

根据我国铝业生产能量流图(图 3-1),计算能源效率 η_e 。能源效率是指铝业生产过程中,一次投入能源最终可以生产出的合格铝产品产量,计算公式为:

$$\eta_e = \frac{1}{E_{r6} + E_n} = \frac{1}{1448 + 922} = \frac{1}{2370} = 0.0422\%$$

同理,计算能源环境效率 q_e 。能源环境效率是指铝业生产过程中,能源流经利用、转换、分配、回收四个系统时对应于单位损失能量的铝产品量,计算公式为:

$$q_e = \frac{1}{E_{r6}} = \frac{1}{1448} = 0.0691\%$$

表 3-3 一般能耗指标

指标	指标计算公式	计算结果
吨铝综合能耗 E_p	$E_p = \sum_{i=1}^n e_i p_i$	21057.4kwh/t
能源效率 η_e	$\eta_e = \frac{1}{E_{r6} + E_n}$	0.0422%
能源环境效率 q_e	$q_e = \frac{1}{E_{r6}}$	0.0691%

由此可知,我国铝业生产企业综合能耗为 2956.46 Kgce/t,还有很大节能空间。在生产细节中注意尽可能降低每一道工序能耗,使能源投入总量得到控制,同时加强回收利用工序产生的余热余能,提高能源效率和能源环境效率,减少能量的损失。

4 基于灰色 GM(1,1)模型的吨铝综合能耗趋势分析

4.1 研究方法与数据来源

本文首先以 2004—2013 年的吨铝综合能耗样本数据为研究基础构建 GM(1,1)模型,然后通过相对误差检验及残差检验对模拟值进行验证,分析模型的精确度,最后利用模型预测 2014—2023 年的吨铝综合能耗。本文研究利用到的数据主要是 2004—2013 年的吨铝综合能耗数据,包括氧化铝综合能耗、铝锭综合交流电耗及铝加工材综合能耗(如表 4-2)。

表 4-2 吨铝综合能耗数据

年份	吨铝综合能耗(Kwh/t)	氧化铝综合能耗(Kgce/t) *	铝锭综合能耗(Kwh/t)
2004	23328.4	1180	15470
2005	23063	1130.5	15294
2006	22517.6	1090.2	14932.3
2007	22310.5	1023.4	14795
2008	21978.8	998.2	14575
2009	22162.8	802.7	14697
2010	21776.7	868.1	14441
2011	21538.4	794.4	14283
2012	21340.9	631.3	14152
2013	21057.4	590.6	13964

* 根据电力当量 0.1404kgce/kWh 折算

由表 4-2 可知,2004—2013 年,吨铝综合能耗表现为逐年持续下降趋势,2004 年为 23328.4kwh/t,2013 年减少到 21057.4kwh/t,10 年期间降低了 2271 kwh/t,年平均降低率为 9.73%。2009 年虽然出现暂时性上升,但 2010 年吨铝综合能耗很快又表现为趋势性下降(如图 4-1 所示)。这说明我国铝工业的能源利用效率正在逐步提高,我国铝业也不断朝着生态化、绿色化铝业发展。

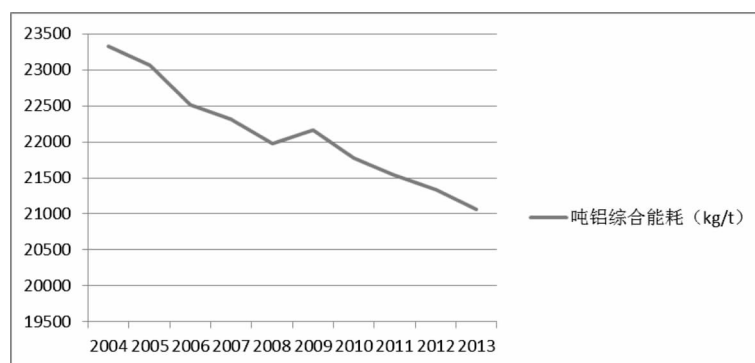


图 4-1 铝综合能耗趋势图

4.2 数据处理与分析

本节对吨铝综合能耗建立灰色 GM(1,1)模型,并根据所建模型对现有科技及经济环境下未来吨铝综合能耗进行科学的预测。为计算方便,统一单位为 kwh/t。

设吨铝综合能耗序列为 $\chi^{(0)}$, 则

$$\begin{aligned}\chi^{(0)} &= (\chi^{(0)}(1), \chi^{(0)}(2), \chi^{(0)}(3), \chi^{(0)}(4), \chi^{(0)}(5), \chi^{(0)}(6), \chi^{(0)}(7), \chi^{(0)}(8), \chi^{(0)}(9), \chi^{(0)}(10)) \\ &= (23328.4, 23063, 22517.6, 22310.5, 21978.8, 22162.8, 21776.7, 21538.4, 21340.9, 21057.4)\end{aligned}$$

第一步:分析验证模型的可行性,进行级比检验。

经计算, $\chi^{(0)}$ 的级比序列为

$$\begin{aligned}\sigma^{(0)} &= (\sigma^{(0)}(1), \sigma^{(0)}(2), \sigma^{(0)}(3), \sigma^{(0)}(4), \sigma^{(0)}(5), \sigma^{(0)}(6), \sigma^{(0)}(7), \sigma^{(0)}(8), \sigma^{(0)}(9)) \\ &= (1.0115, 1.0242, 1.0093, 1.0151, 0.9917, 1.0177, 1.0111, 1.0093, 1.0135)\end{aligned}$$

吨铝综合能耗数据序列 $\chi^{(0)}$ 对应的级比覆盖区域为 $\sigma^{(0)}(k) \in (e^{-\frac{2}{n+1}}, e^{\frac{2}{n+1}}) = (0.8338, 1.1994)$, 所有级比均落在该区域内, 所以可以建立 GM(1,1) 模型。

第二步: 累加一次 $\chi^{(0)}$, 生成序列 $\chi^{(1)}$ 。

$$\begin{aligned}\text{即 } \chi^{(1)} &= (\chi^{(1)}(1), \chi^{(1)}(2), \chi^{(1)}(3), \chi^{(1)}(4), \chi^{(1)}(5), \chi^{(1)}(6), \chi^{(1)}(7), \chi^{(1)}(8), \chi^{(1)}(9), \chi^{(1)}(10)) \\ &= (23328.4, 46391.4, 68909, 91219.5, 113198.3, 135361.1, 157137.8, 178676.2, \\ &200017.1, 221074.5)\end{aligned}$$

第三步: 计算 GM(1,1) 模型的估计参数列 $\hat{\alpha}$ 。

$$\begin{aligned}\text{其中, } B &= \begin{bmatrix} -\frac{1}{2}(\chi^{(1)}(1) + \chi^{(1)}(2)) & 1 \\ -\frac{1}{2}(\chi^{(1)}(2) + \chi^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -\frac{1}{2}(\chi^{(1)}(n-1) + \chi^{(1)}(n)) & 1 \end{bmatrix} = \begin{bmatrix} -34859.9 & 1 \\ -57650.2 & 1 \\ -80064.25 & 1 \\ -102208.9 & 1 \\ -124279.7 & 1 \\ -146249.45 & 1 \\ -167907 & 1 \\ -189346.65 & 1 \\ -210545.8 & 1 \end{bmatrix}, \\ Y_n &= \begin{bmatrix} \chi^{(0)}(2) \\ \chi^{(0)}(3) \\ \vdots \\ \chi^{(0)}(n) \end{bmatrix} = \begin{bmatrix} 23063 \\ 22517.6 \\ 22310.5 \\ 21978.8 \\ 22162.8 \\ 21776.7 \\ 21538.4 \\ 21340.9 \\ 21057.4 \end{bmatrix}\end{aligned}$$

$$\text{经计算, } \hat{\alpha} = (\alpha, \beta)^T = (B^T B)^{-1} B^T Y_n = \begin{bmatrix} 0.010098153 \\ 23220.7192628 \end{bmatrix}$$

其中, $\alpha = 0.010098153, \beta = 23220.7192628$

第四步: 求得时间响应函数, 确定 GM(1,1) 预测模型。

$$(\chi^{(1)}(0) = 23328.4, \frac{\beta}{\alpha} = 2299501.628)$$

GM(1,1) 的时间响应函数为

$$\hat{\chi}^{(1)}(k+1) = \left[\chi^{(1)}(0) - \frac{\beta}{\alpha} \right] e^{\alpha k} + \frac{\beta}{\alpha}, k = 0, 1, 2, \dots, n$$

因此确定 GM(1,1) 预测模型:

$$\left\{ \begin{aligned} \hat{\chi}(k+1) &= -2276173.228 e^{-0.010098153k} + 2299501.628 \\ \hat{\chi}(k+1) &= \hat{\chi}^{(1)}(k+1) - \hat{\chi}^{(1)}(k) \quad (k=0, 1, 2, \dots) \end{aligned} \right\}, \quad (4.15)$$

第五步:GM(1,1)预测模型检验。

(1)残差检验

根据式(4.15),计算得到 $\hat{\chi}^{(1)}(k)$,再累减生成 $\hat{\chi}^{(0)}$ 序列,并计算残差 $\Delta^{(0)}(k)=\chi^{(0)}(k)-\hat{\chi}^{(0)}(k)$,相对残差 $\epsilon(k)=\frac{\Delta^{(0)}(k)}{\hat{\chi}^{(0)}(k)}\times 100\%$,计算结果如表4-3所示。

表4-3 模型实际值、拟合值及残差计算结果

年份	$\chi^{(0)}(k)$	$\hat{\chi}^{(0)}(k)$	$\chi^{(1)}(k)$	$\hat{\chi}^{(1)}(k)$	$\Delta^{(0)}(k)$	$\epsilon(k)\%$
2004	23328.4	23328.4	23328.4	23328.4	9.4587	4.0546
2005	23063	22869.48	46391.4	46197.88	193.5186	0.8390
2006	22517.6	22639.70	68909	68837.59	-122.1040	-0.5422
2007	22310.5	22412.24	91219.5	91249.82	-101.7353	-0.4559
2008	21978.8	22187.05	113198.3	113436.87	-208.2519	-0.9475
2009	22162.8	21964.13	135361.1	135401.01	198.6688	0.8964
2010	21776.7	21743.45	157137.8	157144.45	33.2498	0.1526
2011	21538.4	21524.99	178676.2	178669.44	13.4136	0.0622
2012	21340.9	21308.72	200017.1	199978.16	32.1824	0.1508
2013	21057.4	21094.62	221074.5	221072.78	-37.2216	-0.1767

该模型的平均残差为 $\epsilon_{avg}=\frac{1}{n}\sum_{k=1}^n|\epsilon(k)|=0.4223\%<0.01$,根据表4-1模型的预测精度为一级(优)。

(2)后验差检验

建模序列 $\chi^{(0)}(k)$ 的均值和方差分别为

$$\bar{\chi}=22107.45, S_1^2=528475.4272, S_1=726.963154。$$

残差序列 $\Delta^{(0)}$ 的均值和方差分别为

$$\bar{\Delta}=0.17204, S_2^2=18656.66899, S_2=136.5894176。$$

$$\text{后验差比值 } C=\frac{S_2}{S_1}=0.1867<0.35,$$

$$P=P\{|\Delta^{(0)}(k)-\bar{\Delta}^{(0)}|<0.6745S_1\}=1>0.95$$

经残差检验和后验差检验之后,将检验结果与模型精度检验等级参照表4-1对比,显示本文建立的吨铝综合能耗的GM(1,1)模型的预测精度等级为一级(优)。具体模型拟合程度见图4-2。

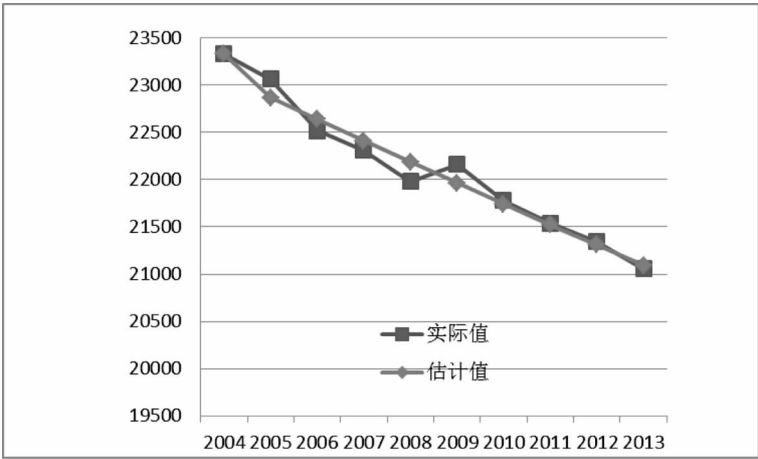


图 4-2 模型计算值与实际值拟合图

图 4-2 表明上述式建立的 $GM(1,1)$ 预测模型的数学表达式具体表示为：

$$\begin{cases} \hat{\chi}^{(1)}(k+1) = 2276173.228e^{-0.010098153k} + 2299501.628 \\ \hat{\chi}^{(0)}(k+1) = \hat{\chi}^{(1)}(k+1) - \hat{\chi}^{(1)}(k) (k=0,1,2,\dots) \end{cases},$$

由此可知，该模型的数据拟合度较好，预测精确度高，与实际数据基本吻合，所以本文能够运用该模型准确预测我国未来几年的吨铝综合能耗。

根据 $GM(1,1)$ 预测公式：

$$\hat{\chi}^{(1)}(k+1) = -2276173.228e^{-0.010098153k} + 2299501.628,$$

分别令 $k=10,11,12,\dots,19$ 。得到未来 10 年吨铝综合能耗的预测值，结果见表 4-4。

表 4-4 未来 10 年吨铝综合能耗预测数据

年份	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
吨铝综合能耗	20883	20673	20466	20259	20056	19854	19655	19458	19262	19068

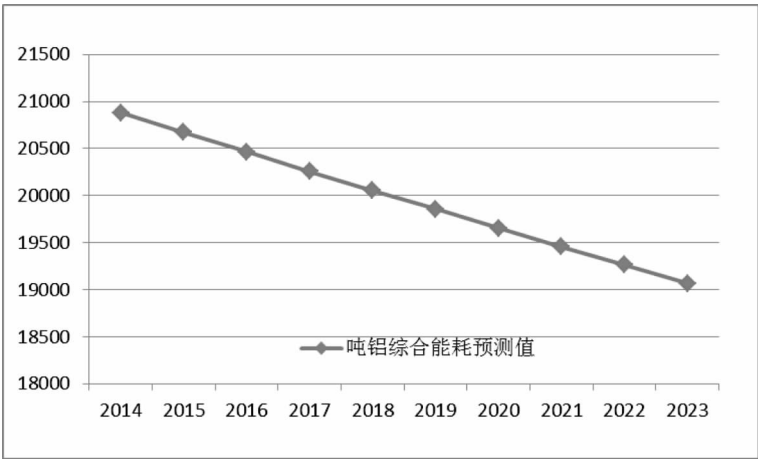


图 4-3 吨铝综合能耗预测趋势

表 4-4 中数据表明,在未来十年内,我国吨铝综合能耗仍然会继续降低,2014 年为 20883kwh/t,2023 年降低至 19068 kwh/t,年均降低率为 8.69%。但与 2004—2013 年的年均降低率 9.73%相比,下降幅度略有减缓,具体变化趋势如图 4-3 所示。

4.3 结论

本文利用我国 2004—2013 年吨铝综合能耗数据,选择 GM(1,1)模型,对未来十年吨铝综合能耗变化趋势进行预测。主要得到以下结论:

(1)利用灰色理论方法实现吨铝综合能耗指标的预测,精度较好,其预测精度等级为一级(优)。从分析结果来看,运用 GM(1,1)模型对历史数据进行分析,其真实值与预测值拟合度较优,预测结果可靠性高,能够反映我国吨铝综合能耗的实际情况和未来发展趋势,这说明运用灰色 GM(1,1)模型预测我国铝业生产吨铝综合能耗的方法切实可行。

(2)GM(1,1)模型与其他数学模型相比,该数学模型简单,对数据的规律性和完整度要求不高,较易建立。但由于 GM(1,1)模型只是根据过去的数据和政策、经济环境展开预测,没有考虑到长期的经济环境、相关国家政策等影响因素的变化,所以 GM(1,1)模型的适用范围局限于短期和中长期预测,做长期预测会影响其准确性。

(3)预测结论表明,虽然未来 10 年我国吨铝综合能耗呈逐年下降趋势,但是我国铝业生产能源问题仍较严峻。因此我国铝业生产应尽快调整改进生产工艺和流程,尽可能减少关键环节的生产能耗。

参 考 文 献

- [1] 王建军,蔡九菊,张琦,吴复忠,陈春霞. 钢铁企业能量流模型化研究[J]. 中国冶金,2006,16(5):48—52
- [2] 闫卫东,孙春强,崔荣国,郭娟. 2015 年全球矿业展望[J]. 中国矿业,2015,01:1—7
- [3] Umeda T, Itoh J, Shiroko K. Heat exchange system synthesis[J]. Chemical Engineering Progress. 1978,74(3): 70—76
- [4] D. W. Townsend, Bodo Linnhof. Heat and power networks in process design. Part I: Criteria for placement of heat engines and heat pumps in process networks[J]. AIChE J, 1983, 29(5): 742—748
- [5] D. W. Townsend, Bodo Linnhoff. Heat and power networks in process design. Part II: Design procedure for equipment selection and process matching[J]. AIChEJ. 1983,29(5):748—771
- [6] R. Smith and B. Linnhoff, The design of separators in the context of overall processes[J]. Chem. Eng. Res. Des. 1988,66(5):196—228
- [7] Joana F. M. F. Cardoso, Johannes IJ. Witte, Henk W. van der Veer. Intra-and interspecies comparison of energy flow in bivalve species in Dutch coastal waters by means of the Dynamic Energy Budget (DEB) theory[J]. Journal of Sea Research. 2006,56:182—197
- [8] Olivier Maury, Blaise Faugeras, Yunne-Jai Shin, etc. Modeling environmental effects on the size-structured energy flow through marine ecosystems Part1: The model[J]. Progress in Oceanography. 2007,74(4):479 - 499
- [9] W. M. Kemp, W. R. Boynton. Productivity, trophic structure and energy flow in the steady-state ecosystems of Silver Springs, Florida[J]. Ecological Modeling, 2004,178(1):43 - 49
- [10] 华贲. 中国能源形势与炼油企业节能问题[J]. 炼油技术与工程,2005,35(4):1—5
- [11] 作浩,刘二恒,华贲. 低温热利用的新格局和系统优化策略[J]. 计算机与应用化学,2009,26(2):133—136
- [12] 吴复忠,蔡九菊,张琦,王建军,陈春霞. 炼铁系统的物质流和能量流的火用分析[J]. 工业加热,2007,36(1):15—18

- [13] Schaltegger S, Synnestvedt T, The link between “green” and between environmental and economic performance [J]. *Journal of Environmental Management*, 2002, 65: 339—346
- [14] 鞠幼化. 冶金企业工序系统节能研究[J]. *节能*, 1994, 6: 8—9

Study on the energy flow in Aluminum production in China

Liu Shishi

(School of Economics and Management, North China Univ. of Tech. , Beijing, 100144)

Abstract The process of Aluminum production is energy-intensive, and rapid development of the Aluminum industry put a great impact on the country's resources, energy and the environment. Therefore, tap the energy potential of Aluminum production and control the emissions become more important problems to the stage of China's Aluminum industry to use energy efficiently and reduce environmental pressures. At first, through in-depth data collection, this paper describes the consumption situation and energy consumption structure on China's aluminum production in detail. Then, the paper constructs the static prospect of the actual energy flow of aluminum production by using the extended three ring section structure model, and accurately estimated the specific energy consumption index. At last, based on Gray GM (1,1) to build predictive models, the prediction accuracy is better, and its true value and the predictive value fit excellently. Highly reliable predictions can reflect the objective reality and development trend of China's comprehensive energy consumption of Aluminum. As a result, GM (1,1) prediction model for comprehensive energy consumption per ton of Aluminum is scientific and rational, it will be able to provide a reliable basis for developing macro energy decisions.

Key words aluminum; energy flow; energy consumption; GM(1,1)

我国铝业生产环境负荷研究

研究生 侯思远

导师 赵贺春

(北方工业大学经济管理学院 北京 100144)

摘 要 铝作为地壳中含量仅次于氧、硅的元素,随着现代化工、农业和科学技术的迅速进步被越来越广泛地运用在各个领域。自建国以来,伴随着工业化进程的不断深入,我国铝业已经走过了六十多年的发展道路。铝因其特殊的属性,对我国航空、建筑、汽车三大重要工业的蓬勃发展起到了十分关键的作用。

随着可持续发展理念的提出,铝业在迅猛发展的同时,其生产过程造成的环境负荷也逐渐引起了人们的重视。无论是铝土矿开采造成的植被破坏、土壤贫瘠,氧化铝生产造成的大量工业三废,还是电解铝生产带来的电力耗费、温室气体排放,都对铝业发展带来了阻力。

为系统研究我国铝业生产的环境负荷问题,本文首先借鉴生命周期理论中清单分析的方法对我国铝业生产一般流程进行了各项环境负荷指标的静态分析,对各生产环节的环境负荷进行排序,明确各环节节能减排的潜力及未来改进方向。之后本文对我国铝业生产的环境库兹涅茨曲线进行研究,分析其是否符合理论中随着经济发展环境污染程度呈现先上升后下降的倒U形形状,并解释其原因。最后借助灰色理论中灰色预测方法,对我国铝业生产的各项环境负荷指标进行未来十年的预测分析,研究在现有条件下我国铝业生产的环境负荷程度能否达到未来铝业发展的要求。本文通过对我国铝业生产环境负荷的静态分析、动态分析及预测分析能够找出我国铝业生产在发展与环境并重的要求下存在的不足及未来的努力方向。

关键词 铝业生产;环境负荷;趋势分析

作者简介 侯思远,女,1989年出生,硕士研究生,主要研究方向为环境会计。

E-mail: eryu11@126.com

1 绪论

铝及其合金制品由于其具有质量轻、耐腐蚀、易延展的优良特性,在航空、建筑、汽车三大重要工业中得到了广泛的应用,产品和消费数量仅排在钢铁之后,是人类使用率最高的金属之一。而据测算,铝的矿藏存储量约占地壳构成物质的8%以上,是资源最丰富的金属元素。由此可见,铝工业不仅在全球工业化进程中起到巨大的推动作用,在未来也将发挥其资源丰富的优势继续成为工业化进程的中坚力量。

与此同时,铝也是生产过程中造成环境负荷最大的金属材料之一。据统计,吨铝能耗为吨钢能耗的约三倍,而电解铝过程中产生的全氟化物,所造成的温室效应更是分别达到了二氧化碳的6500倍和9200倍。

2 我国铝业生产发展现状

2.1 铝土矿开采

铝土矿开采是铝业生产的基础环节,新中国成立以后我国铝土矿开始大规模开采,首先在1954年

恢复了山东沭水矿山。八十年代后,我国的铝土矿开采得到迅速发展,以山西铝厂、中州铝厂为代表的一批大型铝厂应运而生^[1]。2004 年我国超越巴西成为仅次于澳大利亚的全球第二大铝土矿产国^[2]。

我国铝土矿消费一直呈现快速上涨趋势,尤其是 2000 年之后,增势特别明显。进入 21 世纪后,我国铝土矿消费已从 2001 年的 897 万吨增至 8124 万吨,十年增长了 8 倍,年均增速达 24.6%,累计消费量接近 4 亿吨。2011 年,我国首次超过澳大利亚成为全球最大的铝土矿消费国,年消费量占世界铝土矿消费总量的 33.8%。

十九世纪六、七十年代,国有企业自给自足,能够满足我国铝土矿的需求。但到了八十年代,由于耐火材料销售渠道不断拓宽,高铝矿石变得供不应求,使氧化铝生产得到迅猛发展,但也迫使各氧化铝厂片面追求效益,亟需高品位富矿石^[3]。导致自建矿山被迫停产,逐渐形成我国铝土矿以民采为主的局面,据统计,目前民采铝土矿约占我国冶金用铝土矿的 65%。但是民采铝土矿由于监管不严,在开采过程中造成了大量的环境问题。

2.2 氧化铝生产

1954 年山东铝厂的建成投产,标志着我国氧化铝工业的诞生,而当时我国氧化铝的年产量仅为 2 万吨。自 2006 年开始,受国内电解铝生产快速增长,氧化铝需求旺盛的拉动,一批新建和改扩建的氧化铝项目在 2006 到 2007 年期间陆续得到投产,使产量大幅增长,2006 年我国氧化铝产量突破 1000 万吨大关,已连续 5 年成为世界第二大氧化铝生产国。这之后氧化铝产量连年攀升,始终位居世界第一位。

我国氧化铝消费一直保持着快速上涨的态势,2000 至 2011 十余年间,我国氧化铝消费从 620 万吨迅速增长至 3588 万吨,涨幅将近 5 倍,年均增速达到 17.3%。不仅如此,过去 10 年,我国氧化铝消费占全球的比重也一直呈上涨的趋势,截至 2011 年,我国氧化铝消费量占全球氧化铝消费总量的 67.5%,是世界最大的氧化铝消费国^[4]。

除铝土矿资源特性外,导致我国烧结法和联合法在直接能耗和间接能耗等方面都比国外的拜耳法高的原因还在于我国的能源结构。目前,我国的天然气资源利用程度较低,煤、焦炭或重油作为主要燃料,能量转换率和余热利用率较低^[5]。加之我国一些氧化铝生产仍存在规模小、设备技术落后、利用率低下等问题,这些原因共同作用形成了我国铝工业生产率低、能耗高、流程长、成本高的现状。

2.3 电解铝生产

我国第一家电解铝厂是 1952 年开工建设的抚顺铝厂,当时的设计产能仅为 1.5 万吨。1982 年,一些建于六、七十年代的铝厂在国家“优先发展铝”的方针指导下纷纷开始进行改扩建,2000 年以前年均增速为 10%左右,增长较为平稳。2001 至 2003 三年间,产量连续三年增幅超过 20%,接着在 2005 年至 2007 年产量又连续三年增幅超过 20%。之后我国电解铝产量也一直稳步攀升,到 2013 年已突破 2000 万吨大关^[6]。

建国初期到上世纪八十年代我国原铝累计消费 0.28 亿吨,年均增速 9.9%,消费增长较为缓慢。自上世纪九十年代以来,我国原铝消费一直保持着持续上涨的态势。2000 年之后我国原铝消费呈现快速增长势头,消费量由 2001 年的 245 万吨猛增至 2011 年的 1763 万吨,十年间增长了 4 倍多,年均增速更是达到了 17.7%,原铝累计消费量达到 10705 万吨。

碳阳极消耗高、电解槽寿命短一直是我国预焙槽生产存在的短板,而电解铝工业在我国依旧属于高耗能、高污染行业。据统计,电解铝工业总用电量已接近全国电力总消费量的 5%,占整个有色金属工业能耗的 80%。而电解铝生产中排放的二氧化碳、氟化物等温室气体,则是导致全球气候变暖的主要原因^[7]。

3 我国铝业生产环境负荷静态分析研究

借用生命周期评价中清单分析的方法对 2013 年我国铝业生产各个阶段的环境负荷进行静态量化分析。通过对我国铝业生产的资源、能源消耗和污染物排放的状况的静态分析、差异分析、对比分析，希望研究结果可以为降低我国铝业生产总体环境负荷提供可供参考的数据支持，进而促进我国铝业进行清洁生产、节能降耗提供决策依据。

以铝土矿开采的静态研究为例，由于矿山种类、规模与机械化程度不同，因此用于铝土矿开采的能耗也不尽相同。而我国铝土矿因具有品位低、矿层薄、地形复杂等特点，所以与国外相比我国矿石开采的能耗较大。根据矿石开采数据库中的相关数据及参考我国其他学者的研究成果得出铝土矿开采投入产出环境负荷清单，如表 3-1 所示，

表 3-1 铝土矿开采投入产出环境负荷清单

项目		单位	输入输出强度
能源输入	燃料油	kg/t	1.37
	电	kwh/t	16.1
产品	铝土矿	t	1
气态排放物	二氧化碳	kg/t	4.24
	一氧化碳	kg/t	0.00156
	二氧化硫	kg/t	0.00714
	氮氧化物	kg/t	0.0029
	甲烷	kg/t	0.00504
	颗粒物	kg/t	77.2

从铝土矿开采的相关数据来看，这一阶段造成的环境负荷并不明显。但正如绪论中所述，由于我国铝土矿存在品位低的先天不足，又加之开采监管不严的后天缺陷。铝土矿开采阶段造成的环境负荷虽然未能在静态投入产出的数据中体现，但其对我国资源环境造成的巨大破坏是不可恢复性的。

运用清单分析的方法依次对我国铝业生产其他三个阶段进行静态分析，得出造成环境负荷的大小依次为电解铝生产、氧化铝生产、碳素阳极生产、铝土矿开采；通过于世界先进铝业生产环境负荷数据进行对比，得出我国在能源结构改善与温室气体减排等方面仍存在较大潜力。

4 我国铝业生产环境负荷动态分析研究

当一国经济发展处于较低水平时，环境污染程度也较轻。但是随着经济的增长，资源、能源会大量消耗，环境也会日益恶化；而当该国的经济发展到较高水平后，随着经济结构的转变，科技水平的提高，环保意识的增强，环境污染程度逐渐降低，生态环境状况将逐渐得到改善。发展水平与环境之间的这种倒 U 形关系称之为环境库兹涅茨曲线。我国铝业生产发展与其造成的环境负荷之间是否符合倒 U 形关系将在下面进行讨论。

本文选取的经济指标为铝工业的总产值。选取铝土矿的消费量作为环境指标中的资源指标，氧化

铝平均综合能耗是与原铝平均综合交流电耗值作为本研究中的能源指标,氧化铝生产过程中排放的赤泥以及电解铝生产过程中排放的二氧化碳作为本研究中的排放物指标。

下面仅以铝工业总产值和吨铝铝土矿消费量为例进行说明,使用 spss20 软件对铝工业总产值和吨铝铝土矿消费量进行多模型曲线拟合,从各模型拟合结果可以看出,三次函数模型的拟合程度较高。

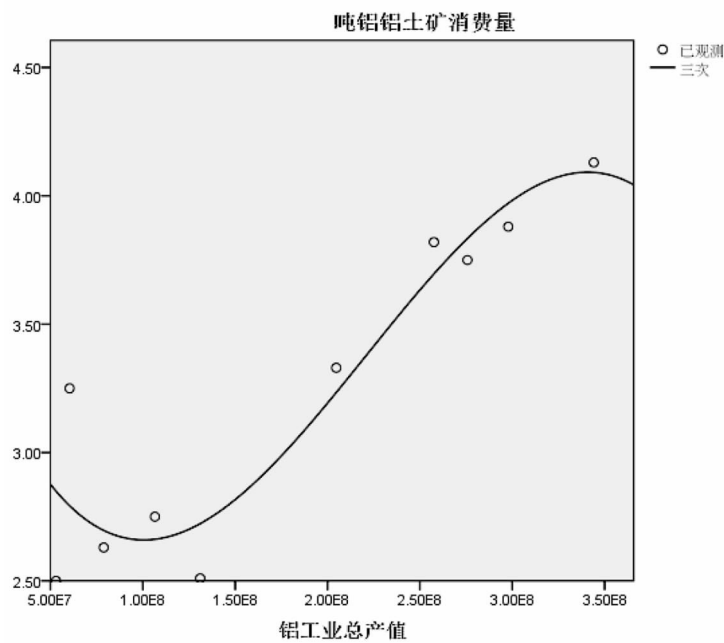


图 4—1 全国铝工业总产值与吨铝铝土矿消费量拟合结果

从拟合图可以看出,全国铝工业总产值与铝土矿消费量的曲线先呈现出正 U 型右侧部分的形状,后呈现出倒 U 型左侧部分形状,随着时间的增长基本呈现持续上升的态势。从发展趋势来看并未出现明显的拐点表明吨铝铝土矿消费量会随着铝工业产值的增加而降低的征兆,也就是说我国目前铝业生产与其造成环境负荷的关系很可能沿着前文所述的将会导致生态环境崩溃的曲线发展。

本研究中选取了铝业生产最具代表性的环境负荷指标对我国铝业生产的环境负荷进行库兹涅茨曲线研究,结果表明我国铝业生产并不符合传统倒“U”型曲线形状,原因在于技术的不普及和产业结构的不合理等因素。

5 我国铝业生产环境负荷预测分析研究

灰色理论是一门应用数学学科,运用于研究信息一部分清楚、一部分不清楚并带有不确定性现象的领域。它是对原始数据进行处理,然后在此基础上建立灰色模型,进而发现并掌握系统发展规律,最终对系统未来状态做出科学预测。我国铝业生产环境负荷研究中,环境负荷指标的数据收集存在很大困难,数据存在滞后性等缺陷,而运用灰色预测模型则能通过现有数据,对未来数年的铝业生产环境负荷进行预测,描绘出我国未来铝业生产环境负荷的大致情况,希望对相关政策的提出与实施和生产水平的提高能够起到一定的辅助作用。

本部分依旧选取了铝业生产典型的五类指标,即吨铝铝土矿消费量、吨铝碳排放量、吨铝赤泥排放

量、氧化铝平均综合能耗、原铝平均综合交流电耗。

将 2004 年—2013 年我国吨铝铝土矿消费量录入灰色预测软件中,将实际量与测量进行对比,得到相关结果,并将结果进行预测精度检验计算,从对我国铝土矿消费量预测结果来看,C 值为 0.43,表明预测精度等级为合格,因此可以使用灰色预测模型对未来十年我国铝土矿消费量进行预测,得到的预测结果如图所示,

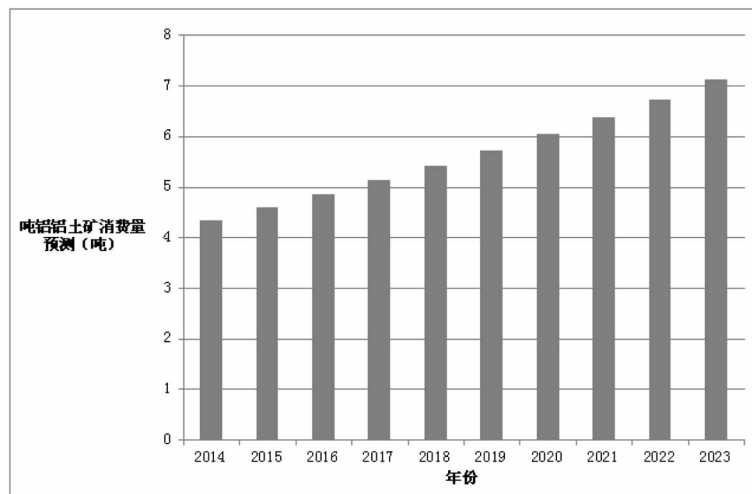


图 5—1 我国铝土矿消费量灰色预测结果

从图 5—1 的预测结果可以明显看出,如果对现在铝土矿消费情况放任不管,低的投入产出比将会大大消耗我国的铝土矿资源,由此带来的植被破坏、水土流失等环境负面影响也将会愈演愈烈。进口国外高品位铝土矿成为解决这一问题的最快方法。但与此同时一些国家限制铝土矿出口的政策也导致我国铝企业必须采用多种方法将铝土矿进口来源多元化,例如在多个国家建立进口铝土矿渠道,在国外建厂换取当地铝土矿进口配额等,避免因某一国家的供应不足导致我国氧化铝生产出现问题。

通过运用灰色理论对我国铝业生产的环境负荷进行预测分析,得出在现有条件下,我国铝业生产中的吨铝碳排放量、原铝平均综合交流电耗等指标在未来不能达到国家可持续发展的要求,仍需在节能降耗等技术上作出提高与改进。

参 考 文 献

- [1] 李昊,王庆飞.我国铝土矿资源开发利用现状及其产业研究[J].陕西科技大学学报:自然科学版,2009,27(5):154—157
- [2] 孙志伟,鹿爱莉.我国铝土矿资源开发利用现状,问题与对策[J].中国矿业,2008,17(5):13—15
- [3] 高峰.生命周期评价研究及其在中国镁工业中的应用[D].北京工业大学学报,2008
- [4] 宋丹娜.基于生命周期评价的铝工业环境负荷研究[D].湖南:中南大学,2007
- [5] 张忠辉.关于四川省电解铝工业“十一五”节能规划的思考[J].四川有色金属,2007(2) 1—4
- [6] 庾莉萍.全球金融危机下的中外铝业[J].轻金属,2009(4):3—7
- [7] 刘杨.中国经济发展与环境污染关系的研究[D].北京:北京工业大学,2010

Study on China's aluminum production environments load

Hou Siyuan

(School of Architecture and Art, North China Univ. of Tech., Beijing, 100144)

Abstract With modern chemical industry, agriculture and the rapid progress of science and technology, as an element in the Earth's crust after oxygen and silicon, Aluminum has been widely used in various fields. Since its founding, along with the industrialization process of China aluminum industry has gone through a development path for more than 60 years. Because of its special properties, aluminum plays a very crucial role in the boom of China's three major aerospace, architectural, automotive industries.

Along with the concept of sustainable development proposed, at the same time of aluminum industry in the rapid development, environmental loads resulting from its production process has also attracted people's attention. Whether due to bauxite mining, destruction of vegetation, soils, alumina production caused by a large number of industrial wastes, electrolytic aluminum production of electricity consumption, greenhouse gas emissions, resistance from the development of the aluminum industry. Our Government in the "Twelve-Five" is definitely proposes to set up a green development plan for building a resource-conserving and environment-friendly society. Indeed, economic development, resource saving, environmental protection, China's aluminum production enterprises have been gradually realized that shoulder social responsibilities, continue to development and improvement of aluminum production technology, to strengthen management and improve efficiency in order to reduce the negative environmental impacts of aluminum production.

For studying China's aluminum production environments load problems, firstly using the theory of life cycle inventory analysis method for reference on China's aluminum production processes on the environment load index of static analysis, to sort the production environments load clearly all links of energy conservation and emission reduction potential and direction for future improvement. Then this paper to study the environmental Kuznets curve of China's aluminum production, analyses their compliance with economic development in the theory of environmental pollution level first increased and then decreased down u-shape, and explain the reasons. Finally, using the grey forecasting method of grey theory and environment load index of aluminum production in China for the next decade of predictive analytics, study on Chinese aluminum industry under current conditions of production environments load aluminum industry can meet the future requirements. Based on production environments load static analysis, dynamic analysis and predictive analytics to identify China's aluminum production at the behest of laying equal stress on development and environmental protection problems and future direction.

Key words aluminum production; environments load; trend analysis

企业技术进步的低碳经济贡献度研究

研究生 魏 洁

导师 王志亮

(北方工业大学经济管理学院 北京 100144)

摘 要 通过收集、整理北京市 2005—2012 年技术进步的关键性指标数据资料,运用灰色系统理论中的灰色关联度模型就技术进步对低碳经济的影响进行综合评价,结论说明北京市技术进步促进二氧化碳减排,也说明北京的低碳经济发展水平在不断提高。为合理确定企业技术进步的低碳经济贡献度,运用模糊层次分析法对北京市低碳经济发展水平进行评价,并通过替代技术进步指标,从而计算出北京市 2005—2012 年技术进步的低碳经济贡献度值。结果显示技术进步对低碳经济的发展是至关重要的,政府部门应大力提倡低碳技术进步,充分挖掘低碳技术节能减排的内在潜力,促进低碳经济的发展。

关键词 低碳经济;技术进步;低碳经济贡献度

作者简介 魏洁,女,1990 年出生,硕士研究生,主要研究方向为环境会计与低碳经济。

E-mail: weijie900921@163.com

1 绪论

全球经济随着工业革命的延续而迅速增长,由此造成了全球碳排放严重超出自然界可以承受的范围等问题,人类也一直在努力为碳减排出“谋”划策。在此背景下,遏制气候变暖,拯救地球家园,使世界各国形成了“同呼吸,共命运”的“责任共同体”。当然,作为迅速发展中的大国,人口基数大,资源相对短缺,经济发展水平较低,同时工业化在不断进步所造成的环境污染等问题,再加上中国现有的碳生产力水平较低,排放一吨二氧化碳所产生的 GDP 仅仅相当于发达国家的 25%左右,这些都显示出我国在碳减排方面的重大压力,我们应当也必须在国际社会上显示我们对低碳发展的决心和信心,提高发展中国家的低碳思想觉悟,加快对低碳排放的投入和发展。

对于“低碳经济”的概念,国内外学者不断充实对它的研究角度和深度,国外著名学者莱斯特·R·布朗(2002)^[1]对低碳经济的实质和内涵进行了早期探索。在布朗对低碳经济的早期探索产生的重大影响下,英国的 Johnston 等(2005)^[2]和德国的 Treffers 等(2005)^[3]分别对其国家的碳减排现状以及如何应对等问题分别进行了研究,认为利用现有技术和相关政策对促进碳减排起到了积极作用,并呼吁相关政府加大对低碳经济的重视,不断深入研究低碳技术的改进和创新。最近 10 多年来,我国学者们紧随外国研究者们对低碳发展的研究脚步,对低碳发展的研究也产生了浓厚的兴趣,有关这方面的研究如奔涌之势涌现出来。如庄贵阳(2005)^[4]、何建坤(2009)^[5]等学者们通过研究发现碳减排的核心在于对新能源技术的研发,通过技术进步与创新可以改善能源利用效率,尽可能的减少碳排放。牛文元(2009)^[6]、贺庆棠(2009)^[7]同样得出,目前实现我国经济低碳化的关键途径是提高化石能源的使用效率,迅速扩大企业在生产和加工过程中对清洁、绿色的可再生能源的使用量,追求一种绿色的可持续发展的生态平衡。

2 企业技术进步的碳减排途径

2.1 促进碳捕集与封存技术

全球变暖问题已经越来越严重,碳捕集与封存技术(简称 CCS)被看作是最具发展前景的解决方案之一。韩文科等(2009)^[8]在研究 CCS 的文章中提到,国际能源署通过长时间的多项研究预测,目前 CCS 是低碳技术中的先进技术,到 2050 年 CCS 在所有限制温室气体浓度的减排技术中占到 20% 的重要地位,并对碳捕集与封存技术进行详细的论述与分析。碳捕集与封存技术是指利用先进的科学技术在二氧化碳排放之前就对其捕集,再由特殊的运输方式将捕集到的 CO₂ 运到封存地,将其压缩注入地下,防止 CO₂ 向上逸散, CCS 技术可分为捕集、运输和封存三个环节。

2.2 提高能源利用效率

能源技术创新的目标主要是使能源的利用率得到提高。比方说,能源开采技术的提高可以最大限度地减少能源开采中的损失,当然技术进步与 R&D 经费和人力资本的投入密不可分。内生增长理论认为,技术进步主要是靠 R&D 经费和人力资本作为支撑的。其中,人力资本为技术进步中的软技术部分,而 R&D 资本则属于硬技术部分,在 GDP 中,R&D 经费所占的比重,可以体现出一个地区或者国家在科技领域投入的多少。

2.3 优化能源消费结构

从增加能源可采储量、促进新能源和可再生能源的增长两方面对技术进步优化能源消费结构进行了说明。我们必须加快资源勘探技术的改进,吸取国内外先进的技术经验,政府部门也要制定相关政策鼓励企业进行资源开发项目的建立,以及对资源勘探方面的资本投入,并指出未来技术进步对新能源和可再生能源开发方面起了至关重要的作用。

3 北京市企业技术进步碳减排效应的量化分析

随着企业技术的不断进步与创新,碳排放强度在逐年随之下降,上文中已经通过灰色关联分析计算出技术进步与碳排放强度之间存在密切的关联关系,那这种关联关系到底有多大?或者说技术进步促进碳排放的效应有多大?为了准确量化企业技术进步对碳减排的促进作用,本文提出企业技术进步的“碳减排效应值”的概念:由于企业技术进步的影响,使得碳排放强度在上年的基础上减少的程度,以此衡量企业技术进步碳减排作用的大小。计算如下:

$$TCC = \frac{CI_0 - CI_1}{CI_0} \times 100\% \quad \text{公式(3.1)}$$

其中, TCC 是指企业技术进步的碳减排效应值; CI 表示为碳排放强度。

根据总结已有的文献研究成果,碳排放的主要影响因素有经济增长、产业结构和技术进步等,为了计算技术进步的碳减排效应值,需要合理的剔除经济增长和产业结构对碳排放的影响,利用碳排放强度指标代表碳排放剔除了经济增长的因素。通过分析各行业之间技术进步碳减排效应的不同,从而有效的排除了产业结构的影响。

表 3-1 2006-2012 年北京市八大行业技术进步的碳减排效应值

行业 \ 年份	2006	2007	2008	2009	2010	2011	2012
农林牧渔业	-8.11%	8.90%	10.79%	10.91%	8.16%	10.29%	14.48%
采矿业	-8.71%	5.06%	7.43%	-44.73%	-1568.22%	83.22%	8.73%
制造业	11.59%	16.82%	22.25%	12.45%	55.57%	7.42%	13.27%
电力、燃气及水的生产和供应业	9.88%	8.20%	-36.46%	13.94%	7.21%	6.93%	8.45%
建筑业	15.30%	8.14%	6.23%	-23.79%	6.58%	23.08%	20.71%
交通运输、仓储和邮政业	-11.33%	-5.66%	-14.44%	7.53%	15.47%	6.87%	-0.67%
批发、零售业和住宿、餐饮业	1.28%	-19.00%	34.19%	8.46%	25.16%	5.98%	4.43%
其他行业	48.76%	-44.30%	11.35%	57.14%	-101.23%	10.93%	9.63%

以各行业各年份技术进步的碳减排效应值为北京市技术进步的碳减排效应值的观测值,将企业技术进步的碳减排效应值变异系数较大的行业赋予较大的权重,技术进步的碳减排效应值变异系数较小的行业赋予较小的比重,从而求得北京市技术进步的碳减排效应值 19.07%。可见企业技术进步在碳减排过程中起到了十分重要的积极作用。

4 北京市企业技术进步的低碳经济贡献度的量化分析

4.1 低碳经济发展水平的评价

为了更为准确的衡量一个地区的低碳经济的发展水平,单一指标很显然不能全面而客观的反映该地区的低碳经济发展水平,最关键的是建立一套普遍被接受的评价指标体系,由于低碳经济是不断发展和进步的,因此也就增加了研究此问题的难度系数,即使有关该项研究的学者们通过一次次不断的改进和完善评价体系,也还是没有形成一套广为使用的低碳经济评价指标体系。本文在大量阅读文献的基础上,通过借鉴学者们的研究体系,遵循低碳经济评价指标体系构建的综合性与系统性、可操作性与可比性的原则,采用分层次和逐级递进的关系来建立了包括总量指标、结构指标和技术指标三大准则层,碳排放总量、三大产业的碳排放比重系数以及技术人员和 R&D 经费比重等 8 项指标在内的一套针对北京市的低碳经济水平评价体系,这套评价体系不仅可以对近几年北京市低碳发展现状进行客观评价,也为度量北京市技术进步的低碳经济贡献度提供基准和依据,为政策制定和城市规划起到一定的指导实践作用。

本文对低碳经济发展水平的评价选取模糊层次分析法(F-AHP),并将通过构建出的低碳经济评价体系计算的综合得分定义为低碳经济发展指数。

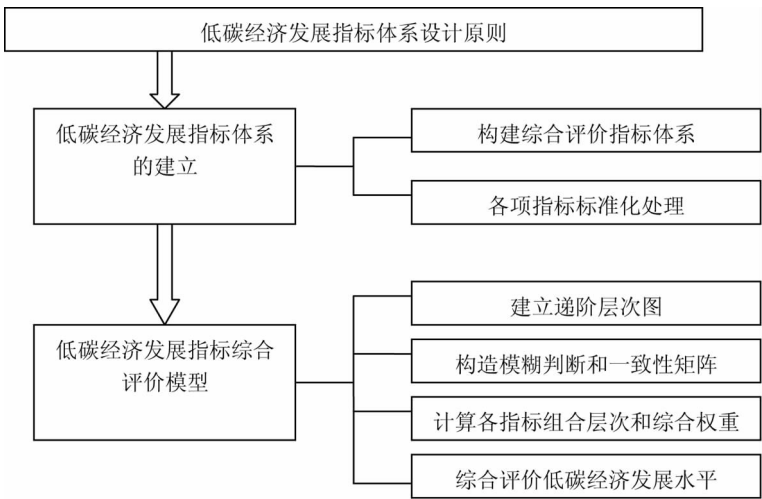


图 4-1 低碳经济发展综合评价整体构思图

(1)构造低碳经济发展指标体系模糊判断矩阵

模糊层次分析的关键就是建立各个指标之间的模糊判断矩阵(优先关系矩阵),用 R 表示模糊判断矩阵,用以表示因素之间的重要性:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$

r_{ij} 表示元素 i 和元素 j 比较时的模糊关系“...比...重要的多”的隶属度,为了定量描述两个于是元素之间的相对重要程度,可以采取如下标度:

表 4-1 模糊判断矩阵(优先关系矩阵)标度

0-1 三标度	含义
0	后者比前者重要
0.5	具有同样重要性
1	前者比后者重要

若 R 满足以下性质:

- 1) $r_{ii}=0.5,(i=1,2,\cdots n)$;
- 2) $r_{ij}=1-r_{ji} ,(i,j=1,2,\cdots n)$;
- 3) $r_{ij}=r_{ik}-r_{jk}+0.5,(i,j,k=1,2,\cdots n)$

则称 R 为模糊一致矩阵。为了减少模糊判断矩阵(优先关系矩阵)的一致性检验,可以事先将优先关系矩阵转化为模糊一致性矩阵,这样就没有必要再进行一致性检验。方法如下:

优先关系矩阵为 R ,假设模糊一致性矩阵为 Q ,且

$$Q = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ q_{21} & q_{22} & \cdots & q_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ q_{n1} & q_{n2} & \cdots & q_{nn} \end{bmatrix}$$

$$q_i = \sum_{j=1}^n r_{ij}, i=1, 2, \dots, n \quad \text{公式(4.1)}$$

$$q_{ij} = \frac{q_i - q_j}{2n} \quad \text{公式(4.2)}$$

(2) 权重计算原理

计算模糊一致矩阵 $Q = (q_{ij})^{n \times n}$ 每行元素的和(不含自身比较)及不含对角线元素的总和:

$$l_i = \sum_{j=1}^n q_{ij} - 0.5, i=1, 2, \dots, n \quad \text{公式(4.3)}$$

$$\sum_i l_i = \frac{n(n-1)}{2} \quad \text{公式(4.4)}$$

因为 l_i 指因素 i 对上层目标的重要性,所以对 l_i 进行归一化处理就可以得到各指标的权重,公式如下:

$$w_i = \frac{l_i}{\sum_i l_i} = \frac{2l_i}{n(n-1)} \quad \text{公式(4.5)}$$

$$U = 0.1235C_{11} + 0.0926C_{12} + 0.0617C_{13} + 0.0926C_{21} + 0.1235C_{22}$$

$$+ 0.0617C_{23} + 0.2222C_{31} + 0.2222C_{32} \quad \text{公式(4.6)}$$

将表 4-2 中的标准化数据代入公式(4.6)可得到北京市近 8 年的低碳经济发展水平的综合得分(见表 4-3)。从表中数据可以看出,北京市有些年份的低碳经济发展指数为负值,但并不是表示其低碳经济发展水平为负。这里的正负仅仅表示该年份的低碳经济发展水平与平均水平的位置关系,将低碳经济发展的平均水平算作零点,这是在运算过程中将数据标准化的结果。根据北京市低碳经济发展综合评价得分,可以对北京市低碳经济发展运行状况进行分析。

表 4-2 各项指标标准化数据

指标 \ 年份	2005	2006	2007	2008	2009	2010	2011	2012
碳排放总量	-0.2027	-0.6373	-1.4519	-0.0713	-0.0759	-0.6965	1.4400	1.6956
人均碳排放量	-1.3002	-1.1045	-0.9904	-0.0945	0.2697	0.4700	1.2633	1.4867
人均生活碳排放量	-1.2649	-1.2649	-1.2649	0.3162	0.9487	0.9487	0.9487	0.6325
第一产业碳排放系数	2.2441	0.2137	0.4461	-0.0141	-0.9811	-0.0992	-1.1386	-0.6708
第二产业碳排放系数	0.1723	-1.0848	-0.1546	-0.2550	-1.6992	0.2904	1.3186	1.4122
第三产业碳排放系数	0.9513	1.4511	0.2511	-0.1620	0.8058	-0.4585	-1.3457	-1.4931
技术人员的比重	-1.2991	-1.2035	-0.8209	-0.8168	0.4194	0.6137	1.1238	1.5382
R&D 经费比重	-1.3960	-1.0758	-0.9957	-0.0548	-0.4054	0.8856	1.0457	1.4060

表 4-3 北京低碳经济发展综合评价指数得分（原始组）

年份\得分	总量指标得分	结构指标得分	技术指标得分	综合评价总得分
2005	-0.2235	0.2878	-0.5989	-0.5345
2006	-0.2590	-0.0247	-0.5065	-0.7901
2007	-0.3491	0.0377	-0.4036	-0.7150
2008	0.0020	-0.0428	-0.1937	-0.2345
2009	0.0741	-0.2510	0.0031	-0.1737
2010	0.0160	-0.0016	0.3331	0.3476
2011	0.3534	-0.0256	0.4821	0.8098
2012	0.3861	0.0202	0.6542	1.0605

根据结果显示,可以分三个阶段对 2005 年—2012 年北京低碳经济发展运行状况进行分析。第一阶段为 2005 年—2009 年。这个阶段北京低碳经济发展综合评价得分小于 0,低碳经济发展比较落后,其中,2005—2008 年技术指标体系得分都小于 0,这说明低碳经济系统对技术进步的投入力度不够,但也可以看出,技术指标得分大体呈现上升趋势,这说明 2005 年—2009 年北京低碳经济发展运行状况虽然不够好,但从技术方面可以看出,北京市对技术方面投入力度在逐年加大;第二阶段为 2010 年。可以称之为过渡阶段,2010 年北京低碳经济发展综合评价得分从负值上升为正值,为 0.3476,由于小于 0.5,可以认为 2010 年为北京低碳经济发展发展状况的第二阶段。2010 年,对于技术指标系统得分大于 0 且比上年上升的幅度明显增大,这说明在这一年,北京政府及企业对低碳技术的重视和努力达到了非常好的效果;第三阶段为 2011 年和 2012 年。北京在 2011 年和 2012 年低碳经济发展指数得分都大于 0.5,相比较 2010 年北京低碳经济发展运行状况来说又上升了一个层次。而且,2012 年的技术指标得分为 2005 年—2012 年中的最大值,这说明北京在低碳技术的改造方面投入越来越大,政府对低碳技术方面的重视程度也由此可见,同时验证了技术进步对低碳经济发展的贡献是很大的。总的来讲,北京市 2005—2012 年的低碳经济发展水平整体呈上升趋势,也是对近几年北京市政府和企业对低碳经济发展所作出的努力的一种肯定,更增加了我们对发展低碳经济的信心,为进一步计算技术进步的低碳经济贡献度奠定了坚实基础。

4.2 北京市企业技术进步的低碳经济贡献度的计算

为了进一步分析在低碳经济发展系统中技术进步对低碳经济发展的影响程度,计算出技术进步的低碳经济贡献度值的大小,本节将技术指标层次的技术研发人员占总从业人数比重和 R&D 经费占地区生产总值比重数据进行逐年替代,即用 2006 年的技术进步指标数据替代 2005 年低碳经济发展综合评价的公式中的技术进步指标数据,下表 4-4 所示是替代后的各项系统的得分情况。

表 4-4 北京低碳经济发展综合评价分数(替代组)

年份	得分	总量指标 得分	结构指标 得分	技术指标 得分	综合评价 总得分
2005		-0.2235	0.2878	-0.5065	-0.4421
2006		-0.2590	-0.0247	-0.4036	-0.6873
2007		-0.3491	0.0377	-0.1937	-0.5050
2008		0.0020	-0.0428	0.0031	-0.0898
2009		0.0741	-0.2510	0.3331	-0.0516
2010		0.0160	-0.0016	0.4821	0.4623
2011		0.3534	-0.0256	0.6542	0.9819
2012		0.3861	0.0202	0.6542	1.0605

根据表 4-3 和表 4-4 中的技术指标得分可以绘制如图 4-2 所示的原始组和替代组的技术进步指标得分的对比柱状图。

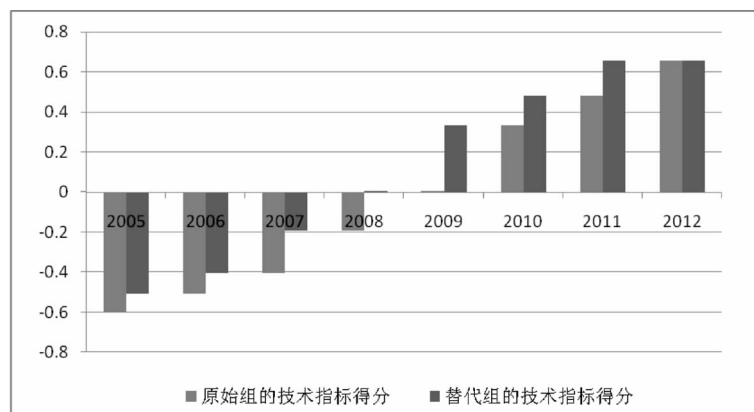


图 4-2 北京市技术进步指标得分对比柱状图

经过上文的理论分析及计算结果可以看出,替代组的低碳经济发展综合评价得分在替代之后总体下降了,说明技术进步对低碳经济的发展具有正向的促进作用。以替代组的指标系统综合得分为基础,本文定义技术进步对低碳经济发展水平提高的程度为技术进步的低碳经济贡献度,用符号 T 表示,计算公式如下:

$$T = \frac{D_{\text{替代组}} - D_{\text{原始组}}}{D_{\text{原始组}}} \times 100\%$$

其中, T 为技术进步的低碳经济贡献度; $D_{\text{原始组}}$ 为原始组数据计算的北京市低碳经济发展综合评价得分; $D_{\text{替代组}}$ 为替代组数据计算的北京市低碳经济发展综合评价得分。

由此,我们能够计算出北京市不同年份的技术进步对低碳经济发展的贡献度,数据如下表所示:

表 4—5 北京市技术进步的低碳经济贡献度

年份	2006	2007	2008	2009	2010	2011	2012
贡献度	17.29%	13.01%	29.37%	68.68%	86.48%	42.84%	21.26%

以 2005—2012 年北京市的低碳经济发展水平为观测值,计算得出北京市企业各年份技术进步的低碳经济贡献度。从整体上分析,企业各年份技术进步的低碳经济贡献度值都大于零,量化分析出技术进步对低碳经济的发展是正向的促进作用,随着企业对技术进步的两项指标(技术研发人员和 R&D 经费)投入力度的增加,技术进步对低碳经济发展的促进作用越明显。

5 结语

随着全球气候的不断变暖,发展低碳经济已经成为世界各国的必然选择,我国政府在“十二五”规划中首次将促进二氧化碳减排列入头号重要任务,再次表明了中国发展低碳经济的决心。怎样发展低碳经济,怎样促进碳减排,成为了各国学者的主要研究方向。本文在大量文献阅读的基础上,通过归纳分析国内外学者的研究成果,认为技术进步是提高能源利用效率,减少单位产出的能源投入,优化能源消费结构的重要途径,且技术进步对低碳经济的发展具有显著的促进作用。通过计算北京市目标年份的技术进步的低碳经济贡献度可以看出,技术进步对发展低碳经济具有显著的促进作用,由此可以说明企业在大力提倡技术进步的同时能够促进低碳经济的同向发展,政府部门应重视对技术进步的政策制定,促进低碳技术的创新和充分利用技术进步来促进低碳经济的发展。

参 考 文 献

- [1] [美] 莱斯特·R·布朗. 生态经济:有利于地球的经济构想[M]. 台湾:东方出版社,2002
- [2] Johnston D, Lowe R, Bell M. An exploration of the technical feasibility of achieving CO₂ emission reductions in excess of 80% within the UK housing stock by the year 2050[J]. Energy Policy, 2005, (33):1643—1659
- [3] Treffers T, Faaij, Sparkman J. Exploring the possibilities for setting up sustainable energy systems for the long-term: two visions for the ditch energy system in 2050[J]. Energy Policy, 2005, (33):1723—1743
- [4] 庄贵阳. 中国经济低碳发展途径与潜力分析[J]. 国际技术经济研究, 2005, 8(3):8—12
- [5] 何建坤. 发展低碳经济关键在于低碳技术创新[J]. 绿叶, 2009, (1):46—50
- [6] 牛文元. 低碳经济是落实科学发展观的重要突破口[N]. 中国报道, 2009—3—19(3)
- [7] 贺庆棠. 低碳经济是绿色生态经济[N]. 中国绿色时报, 2009—8—4(2)
- [8] 韩文科, 杨玉峰. 全球碳捕集与封存技术的最新进展[J]. 宏观经济研究, 2009, (12): 22—24

A Study on the Contribution of Technological Progress to the Low Carbon Economy

Wei Jie

(School of Economics and Management, North China Univ. of Tech., Beijing, 100144)

Abstract By collecting and sorting the data of 2005—2012 about technological progress of Beijing, established the comprehensive evaluation indexes of the impact of technological progress on the low-carbon economy by using gray correlation model of gray system theory, which indicated that technological progress can promote carbon dioxide emissions and also showed that Beijing's low carbon economy development level is improving. In order to get a reasonable evaluation of the development level of Low Carbon Economy, the paper used F—AHP to calculated the comprehensive scores of low-carbon economy system. In order to get a reasonable conclusion about the Contribution of Technological Progress to the Low Carbon Economy, we used the technical indicators of this year replaced by the next year, which calculated the finally contribution to Beijing's low carbon economy from 2005 to 2012. The results show that technical progress is crucial for the development of low-carbon economy, the government departments should vigorously promote low-carbon technological progress, and fully exploit the inherent potential of Energy conservation about low-carbon technologies to promote the development of low-carbon economy.

Key words low carbon economy; technological progress; contribution of low carbon economy

基于商业生态系统的产业链协同创新机理研究

——以新能源汽车为例

研究生 李雅荣

导师 吴永林 纪雪洪

(北方工业大学经济管理学院 北京 100144)

摘 要 近几年,新能源汽车的市场表现势头良好,涌现出一批从事新能源汽车业务的企业,其中特斯拉汽车、比亚迪和 Better Place 三者的表现最让人值得思考。针对此问题的深入探讨,本文采用多案例研究与扎根理论相结合的研究方法。文中运用扎根理论,对所收集资料进行三级编码,得出初始理论:产业链协同创新的关键节点有四步——创新策略、信任合作、价值联盟以及创新共生。运用多案例分析对所得初始理论进行修正、丰富和完善。从而得出在商业生态系统中,产业链进行协同创新的机理过程—首先要找到合适的创新策略,这是创新的基础;更为重要的是,在协同创新过程中,企业之间要建立起优良的信任合作关系和稳固长性的价值联盟;最终促使各企业实现创新共生。

关 键 词 协同创新; 创新策略; 信任合作; 价值联盟; 创新共生

作者简介 李雅荣,女,1991 年出生,硕士研究生,主要研究方向为战略管理。

E-mail: yarong0203@163.com

1 绪论

随着经济全球化的深入发展,当代企业正面临着越发复杂多变的外部环境。在经济发展大浪潮中,企业抓住机遇可以获得新的发展;相反,如果错失发展良机,则会给企业带来很多意想不到的风险。因此企业的发展不能再仅仅局限于自身的发展,更要关注所在生态系统的发展。市场竞争的加剧、环境污染日益严重以及气候的变化,传统汽车行业的发展了受到诸多严峻的挑战。随之而来,越来越多的目光开始关注于新能源汽车的发展。经过近几年的发展,取得了许多成就,但问题同样存在,其中商业生态系统中企业间的协同创新能力的构建问题尤为突出,由此从商业生态系统入手研究全球新能源汽车产业的发展,有其一定的必要性和重要性。本文从商业生态系统的角度入手,将产业链上与核心企业的成长利益攸关的企业或机构都纳入考量范围,重点探究企业间的协同创新活动,以及通过相互之间的影响提升整个商业生态系统的协同创新能力的过程和机理机制。

2 研究设计

首先,在对文献研究的基础上,本文回顾了国内外学者对商业生态系统^[1]及协同创新^[2]问题的相关研究,对其主要理论和观点^[3]进行了整理,发现目前的研究主要集中于协同创新^[4]活动的影响因素^[5]及其评价^[6]、绩效评价体系的建立^[7]等,而缺少对具体的协同创新活动的关键行为的研究,因此本文力图在商业生态系统理论的基础上,找出产业链进行协同创新的关键行为节点。

随后,本文对研究设计进行了详细的说明,科学的研究设计能够有效提高案例研究结论的信度和

效度。因此,从研究方法、案例选择、数据收集以及数据分析策略四个方面对本项研究的科学性进行了论述。本文确立了多案例研究与扎根理论相结合的分析方法,通过五项案例选择指标,在新能源汽车行业中的多家企业中选择出较有代表性的三家典型企业,采用合理的多样的方式来收集相关的一手资料和二手数据,在数据和资料的分类整理和深层次加工处理上运用了扎根理论这种研究方法,对数据进行三级编码,尝试建立初始理论,并深入探讨该商业生态系统中企业间协同创新能力的构建问题和商业生态系统中产业链的协同创新过程。

3 数据分析过程

扎根理论是一种定性研究方法,主要就是在经验资料的基础之上建立理论。本文展现了基于扎根理论的数据分析全过程,文中运用扎根理论的方法,对所收集的全部资料进行三级编码:开放式编码、主轴式编码以及选择式编码。

3.1 开放式编码

对于各电动汽车厂商来讲,实现电动汽车的市场化过程中最大的难题就是解决电动车的续航能力并提高消费者和市场的接受度和认可度。针对这一难题,文中所研究的三家企业分别选取不同的角度为切入点来进行解决,从而采取了不同的商业行动。

通过严密的开放式编码过程,在三家企业中出现的个案资料和现象中提炼概念,并在所得概念的基础上进行归纳,从而得到本案例研究的19个初始范畴——生态环境、组织资源、创新路径、系统资源、外部性、合作经验、合作态度、核心能力、知识流动性、市场严峻、利益同盟、成员认可度、产品竞争能力、创新收益专属权、组织间的溢出效应、共同愿景、学习能力、技术开发能力、网络的稳定性^[8]。

3.2 主轴式编码

主轴式编码阶段借助了译码典范,分析初始范畴之间的存在的因果条件、脉络、中介条件、行动/互动策略以及结果^[9],对初始范畴进行相应的主轴式编码,探讨初始范畴之间内含的逻辑关系。

借助译码典范,我们可发现在初始范畴生态环境、组织资源、创新路径、系统资源中内含的逻辑关系为:企业的创新策略制定依据就是根据所在的生态环境,充分考量已有的组织资源,从而寻求相应的创新路径,在这个过程中,系统资源起着一定的中介作用,因此将这四个初始范畴归纳为第一个主范畴——创新策略。

对于初始范畴外部性、合作经验、合作态度、核心能力、知识流动性,文中得出其内含的逻辑关系为:出于资源互补和经济外部性的考量,在合作经验的基础上,建立长久的信任合作关系,合作态度和核心能力会起着一定的中介作用,加快系统的知识流动性,因此将这五个初始范畴归纳为第二个主范畴——信任合作。

对于初始范畴市场严峻、利益同盟、成员认可度、产品竞争能力、创新收益专属权,文中得出其内含的逻辑关系为:面临严峻的市场形势,在利益关系的驱动下,建立系统内的价值联盟,成员认可度和产品竞争能力有着一定的中介作用,最终形成核心企业的创新收益专属权,因此将这四个初始范畴归纳为第三个主范畴——价值联盟。

对于初始范畴组织间的溢出效应、共同愿景、学习能力、技术开发能力、网络的稳定性,文中发现其内含的逻辑关系为:基于组织间的溢出效应,促进系统实现创新共生的理想状态,其中学习能力和技术开发能力有着一定的中介作用,最终保证了网络的稳定性,因此将这四个初始范畴归纳为第四个主范畴——创新共生。

3.3 选择式编码

得到四个主范畴之后,开始进行选择式编码。挑选和总结在对主轴式的编码阶段中得到的主范畴之间的逻辑关系,形成一条“故事线”,提炼出核心范畴。所提炼出的核心范畴可以被系统地识别和关联使得选择式编码概念的形成,核心范畴的作用是使得其他相关范畴要围绕它结合起来。

因此,在总结主范畴之间的逻辑关系的前提下,本文开发出一条合理的“故事线”:在一个商业生态系统之中,每个企业会根据其生态环境和技术基础找到符合自身的创新策略,基于这样的创新策略,企业寻找相应的合作伙伴,即合适的供应商、市场、金融机构等,在合作的过程中构建利益共享的价值联盟,进而实现整个生态圈的创新共生,形成所处生态系统的协同创新能力。从而提炼出本文的核心范畴:以商业生态系统为基础的产业链协同创新。

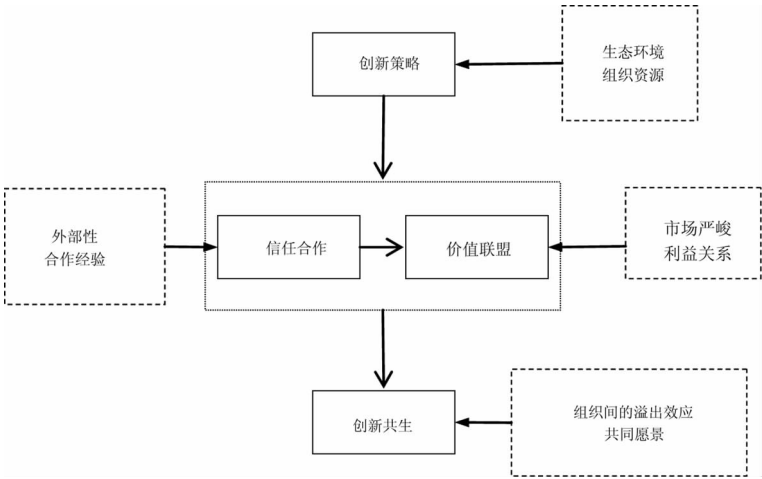


图 3—1 商业生态系统产业链协同创新的编码结果

4 研究发现

4.1 案例分析结果

针对扎根理论阶段得出的初步理论,本文采用多案例进行横向对比以及反向论证,对初始理论进行相应的修正、丰富和完善。具体的案例分析结果如下:

表 4—1 案例分析结果示表

	比亚迪	特斯拉	Better Place
创新策略	坚定地走技术研发路线	理念创新+技术研发	主张换电模式
信任合作	垂直整合产业链,并持续的向上下游延伸	在全球范围对产业链进行外包整合	可合作车企较少 合作关系单薄
价值联盟	吸纳多方成员 联盟程度较高	吸纳多方成员 联盟程度较高	供应商单一 联盟程度较低
创新共生	共生状态良好,能够实现持续的技术创新	共生状态良好,能够实现持续的技术创新	无法实现持续的技术创新

	比亚迪	特斯拉	Better Place
协同创新能力	各参与企业可以有秩序的进 行协同创新活动	各参与企业可以有秩序的进 行协同创新活动	协同创新活动较为混乱
市场表现	良好	良好	走入破产保护程序

4.2 主要观点

基于上述案例对比分析过程,得出论文的主要观点:

(1)协同创新能力的影响因素包括知识流动性、创新收益专属权、网络的稳定性;(2)作为核心企业,要慎重选择其进入新市场时的创新策略,这会对其生态系统的演变和发展有着深远影响;(3)生态系统内要保证较高度的信任合作,这是保证知识流动性的基础,也是系统提高协同创新能力的重要保证;(4)生态系统内要构建稳健的价值联盟,这是形成创新收益专属权的前提,也是系统提高协同创新能力的重要保障;(5)生态系统内要实现企业间的创新共生,这可保证系统的稳定性。

由此得到对初始理论修正后的概念模型,创新策略持续影响生态系统的协同创新过程,信任合作、价值联盟、创新共生通过知识的流动性、创新收益专属权、网络的稳定性对系统的协同创新能力产生作用,最终影响整个生态系统的协同创新活动。具体如下图所示:

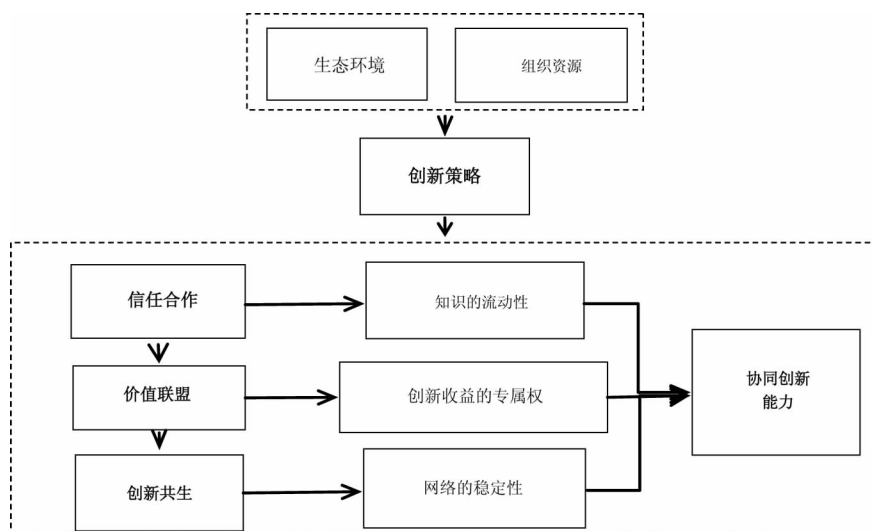


图 4-1 商业生态系统中的协同创新概念模型

5 结论

基于上述分析,本文最终得出在商业生态系统中产业链协同创新的过程就是系统内企业共同提升协同创新能力的过程。以核心企业的创新策略为整体出发点,通过相互之间的信任合作促进系统内知识和信息的流动,构建共同认可的价值联盟确保创新收益的专属权,最终实现整个商业生态系统的创新共生,保证了系统的持续稳定发展。具体主要包括以下五个方面的协同过程:

(1)作为核心企业,其进入新市场时的创新策略选择对其生态系统的演变和发展有着深远影响。具体来讲,企业的创新策略很大程度上决定了企业的创新资源配置以及合作伙伴的选择,这对于整个

生态系统的建立和完善都有着最直接的关系。

(2)生态系统内的信任合作是保证知识流动性的基础,是系统提高协同创新能力的重要保证。系统内信任机制的完善程度直接影响信息的传递效果以及知识的流动程度,这对于一个需要创新积累和创新突破的生态系统至关重要。

(3)生态系统内价值联盟是形成创新收益专属权的前提,系统提高协同创新能力的重要保障。创新收益的专属权反映的是所属企业拥有的排它技术能力,可使企业获得在市场竞争优势的制胜法宝,这就要求核心企业尽可能建立和完善它的价值联盟,提高整个系统的协同创新能力,发挥其功能。

(4)生态系统的创新共生保证了网络的稳定性,是生态系统形成协同创新能力的最佳状态。系统稳定才能获得发展,所有的生态系统发展的最终目的必将是实现系统的有竞争性的稳定状态,因此,企业间的创新共生成为了系统成员的共同目的。

(5)商业生态系统中产业链协同创新的过程就是一个系统内企业共同提升协同创新能力的过程,通过信任合作促进系统内知识和信息的流动,构建共同认可的价值联盟确保创新收益的专属权,最终实现系统的创新共生,确保了整个系统的持续稳定发展。

本文从商业生态系统视角来考量新能源汽车产业链上的企业,完善了核心主体的利益相关者,从商业生态系统的角度完善了产业链上的企业进行协同创新活动的过程和方式,细化商业生态系统中产业链上企业进行协同创新活动的内在机理。

近年来,新能源汽车的发展给我们带来了新的机遇。但由于关键技术的瓶颈限制、产业链的上下游企业不能很好地进行协同合作,使得新能源汽车的发展严重受限。因此,本文对我国企业产业链的协同创新活动进行了详细的多案例分析,深入研究了新能源汽车产业链中的企业进行协同创新活动的内在机理,明确产业链企业在协同创新活动中的关键行为节点,为企业进行协同创新提供了较为明确的行动指南。整篇论文的研究结果以期对我国新能源汽车产业的持续发展提供一系列完整的指导意见,从而促进我国新能源汽车产业的稳定长久发展。

参 考 文 献

- [1] James F Moore. Predators and prey:a new ecology of competition [J]. Harvard Business Review,1993,(5):75—86
- [2] 陈劲,王方瑞. 再论企业技术和市场的协同创新——基于协同学序参量概念的创新管理理论研究[J]. 大连理工大学学报,2005,(6):1—5
- [3] Marco L ,Roy L. The Keystone Advantage:what the new dynamics of business ecosystem mean for strategy,innovation and sustainability[M]. Harvard Business School publishing Corporation, 2004. 25—49
- [4] 洪柳. 商业生态系统内企业间协同问题研究[J]. 商业经济研究,2010,(8):21—22
- [5] Persaud A,Kumar U,Kumar V. Coordination Structures and Innovative Performance in Global R&D Labs[J]. Canadian Journal of Administrative Sciences,2002,(1): 57—75
- [6] 杨丽伟. 产业链企业协同创新的内部影响因素研究[J]. 中国市场,2011,(15):55—56
- [7] 高晓霞. 基于产业集群的高技术企业协同创新绩效研究[D]. 太原:太原理工大学,2014
- [8] 范太胜. 基于产业集群创新网络的协同创新机制研究[J]. 中国科技论坛,2008,(7):26—30
- [9] 王丽华. 基于创新集聚的协同创新机理研究[D]. 大连:东北财经大学,2012

Research on the Collaborative Innovation of Industry Chain Based on Business Ecosystem:a case study of new energy vehicle industry

Li Yarong

(School of Economics and Management, North China Univ. of Tech. , Beijing, 100144)

Abstract In recent years, new energy automobiles achieve the booming development, a group of companies spring up, including Tesla Motors, BYD and Better Place, whose performance worth us thinking about. In order to find out the reasons of their success or failure, this paper combines research methods of case study and grounded theory. Using the grounded theory, the collected data are three-level coding. We find out the initial theory: the collaborative innovation in a industry have four key steps - innovation strategy, trust and cooperation and value alliance and innovation symbiosis. Then use the cases analysis to modify, enrich and perfect the initial theory. So we find out the mechanism of collaborative innovation process in the business ecosystem. First, find the right innovation strategy, which is the basis of innovation; More importantly, in the process of collaborative innovation, enterprises should establish good trust and cooperation and firm value alliance; Eventually , promote various enterprises to realize innovation symbiosis.

Key words business ecosystem; collaborative innovation; innovation strategy; cooperation and trust; the value alliance; innovation symbiosis

文化创意企业团队异质性、 知识共享与团队绩效关系研究

研究生 秦 爽

导师 郑强国

(北方工业大学经济管理学院 北京 100144)

摘 要 本文在文献研究的基础上,根据文化创意企业团队特征,引入知识共享作为中介变量,建立了团队异质性、知识共享与团队绩效的理论模型,并以 130 个文化创意团队为样本,利用 SPSS19.0 软件及 AMOS17.0 结构方程模型软件,验证了团队异质性、知识共享与团队绩效的作用机理。研究表明:社会分类异质性对团队绩效和知识共享均起到了负向作用;信息异质性对团队绩效和知识共享均起到正向作用;价值观异质性对知识共享和团队绩效中的创新绩效、成员满意度均起到负向作用,而对任务绩效的负向作用未达到显著水平;知识共享在团队异质性与团队绩效间起到了中介作用。最后结合研究结果,为文化创意企业团队管理提出了建议。

关 键 词 文化创意企业;团队异质性;知识共享;团队绩效

作者简介 秦爽,女,1990 年出生,硕士研究生,主要研究方向为组织与人力资源管理。

E-mail: qinshuang0814@163.com

1 绪论

异质性团队是文化创意企业最主要的工作单元,成员不仅在知识、技能层面存在差异,同时在性别、年龄及价值观层面也存在差异。知识共享是文化创意团队非常重要的团队过程,也是影响团队创造力提升的关键,但现有研究中鲜有学者探究知识共享在团队异质性与团队绩效间的中介作用。本文根据文化创意团队独特的工作特点,提出具有针对性的中介变量——知识共享,弥补了现有研究中的不足,具有良好的理论和实践价值。

2 研究综述与假设提出

2.1 概念界定

团队异质性是使个体感知到自己与他人不同的各种差别^[1],如性别、年龄、种族、专业、价值观等。本文借鉴 Jehn(1999)的思想将团队异质性分为了社会分类异质性(年龄、性别与团队任期)、信息异质性(学历、专长与职能部门经历)和价值观异质性三个维度^[2]。本文聚焦于团队内部的知识共享,因此将知识共享定义为团队成员彼此利用各种渠道交换知识并创造新知识的过程。本文的团队绩效指团队在任务相关、团队创新及成员满意度三个方面达成的效果。

2.2 团队异质性对团队绩效的影响

社会认同理论和相似吸引理论表明,具有相同社会分类的个体之间更容易产生认同,形成非正式组织。学者们的实证研究也证实了团队成员在年龄、性别和团队任期等社会分类变量上的异质性会引

起人际冲突,破坏团队凝聚力,从而降低团队绩效^[3]。信息决策理论表明,异质性不仅为团队提供了差异性知识资源,同时不同的教育背景与生活经历为团队处理问题提供了迥异的方法与角度,有助于团队绩效的提升。文创团队成员具有较强的主观意识与极高的个性化与自我实现的追求^[4],价值观层面的异质性会引起团队成员间的争执,破坏团队合作氛围,降低团队绩效。

综上,本文提出如下研究假设:

H1:团队异质性与团队绩效显著相关

H1a:社会分类异质性对团队绩效有显著负向作用

H1b:信息异质性对团队绩效有显著正向作用

H1c:价值观异质性对团队绩效有显著负向作用

2.3 团队异质性对知识共享的影响

研究发现,在我国这种高关系文化背景下,关系距离是员工在面临知识共享决策时考虑的首要因素^[5]。根据社会分类理论,异质性团队成员在社会分类及价值观层面的异质性会加剧知识贡献者与知识收集者的关系距离,降低知识贡献者及知识收集者双方的知识转移意愿,从而阻碍个体间的知识共享。然而影响团队成员知识共享的因素除了共享意愿外,还有知识因素。李夏楠和倪旭东(2012)对知识差异程度与知识共享度的研究表明,在知识差异度低的团队中知识共享度较低,随着知识差异度的提高,团队知识共享度也随之提高^[6]。

综上,本文提出如下研究假设:

H2:团队异质性与团队知识共享显著相关

H2a:社会分类异质性对知识共享有显著负向作用

H2b:信息异质性对知识共享有显著正向作用

H2c:价值观异质性对知识共享有显著负向作用

2.4 知识共享对团队绩效的影响

创新来自异质性知识间的碰撞与整合运用,知识共享可以拓展创新思维,引发创意产生,促进团队任务绩效和创新绩效的提升^[7]。成员满意度是员工对自己的工作和组织的满意程度。创意人才对社会尊重和自我价值实现具有极高需求,同时文化创意产品的设计与开发需要融入大量文化、美学和艺术知识^[8]。知识共享可以提升个人在团队中被认同与被尊重感知,实现个人隐性知识间的传递,有助于激发创意人员的创作灵感,提升其工作满意度。

综上,本文提出如下研究假设:

H3:团队知识共享对团队绩效有显著正向作用

H3a:知识共享对团队任务绩效有显著正向作用

H3b:知识共享对团队创新绩效有显著正向作用

H3c:知识共享对团队成员满意度有显著正向作用

2.5 知识共享的中介作用

通过前面的分析可知,在文创企业中,知识共享是团队中重要的团队运作过程,不同维度的团队异质性会从不同角度对知识共享产生影响,而知识共享又是提高团队绩效的重要因素,因此本文进一步提出如下研究假设:

H4:知识共享在团队异质性与团队绩效间起中介作用

通过对文化创意团队异质性、知识共享与团队绩效间关系的梳理,本文构建出如下的理论模型。如图 2-1 所示:

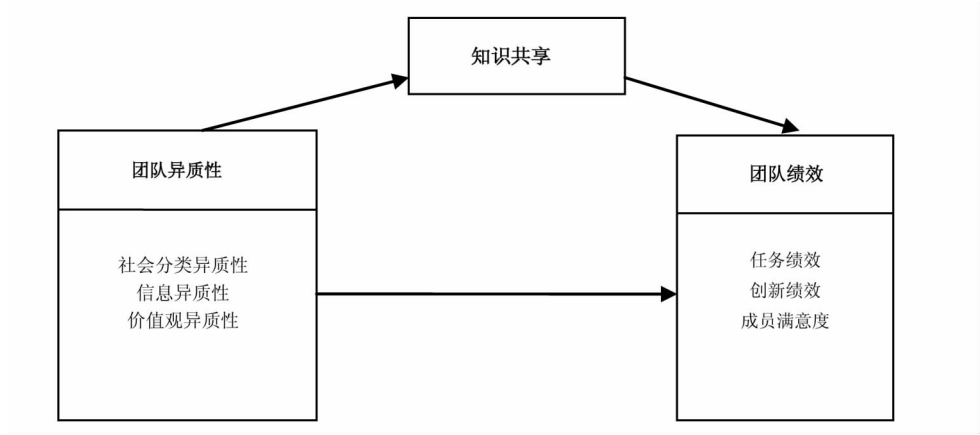


图 2-1 团队异质性、知识共享与团队绩效关系理论模型

3 问卷设计与收集

3.1 变量测量

本文社会分类异质性和信息异质性根据团队成员特征分布情况,采用 Blua(1997)指数进行测量,计算公式为 $H=1-\sum P_i^2$,其中 i 为变量分类总数, P_i 为第 i 个类别的成员在团队中所占的比例。 H 值介于 0—1 之间,越接近 1 表示团队异质性程度越高。价值观异质性、知识共享和团队绩效(任务绩效、创新绩效和成员满意度)变量均采用李克特五级量表进行测量。各变量测量量表均在综合分析其文化背景、适用性、量表信、效度的基础上,翻译修订自国内外成熟量表。

3.2 数据收集与样本结构

本次调研以团队为单位,针对北京地区的 169 个文化创意团队发放了调查问卷 724 份,成套回收问卷 142 个团队,剔除无效问卷后,有效团队数 130 个,共计 607 份有效问卷,覆盖了 61 个文化创意企业,团队成套有效回收率为 76.9%。样本特征如表 3-1 所示:

表 3-1 样本特征描述 (n=130)

项 目		百分比%	项 目		百分比%
团队规模	2—5 人	56.9	行业	文化艺术	6.2
	6—10 人	19.2		新闻出版	5.4
	11—15 人	13.8		广播、电视、电影	6.2
	16—20 人	4.6		软件、网络及计算机服务	30.0
	20 人以上	5.4		广告与会展	15.4
				设计与服务	32.3
				其他	3.1
团队组建时间	3 个月及以下	16.9	企业规模	50 人以下	20.8
	3 个月—1 年	26.9		50—100 人	16.9
	1—3 年	33.1		101—200 人	14.6
	3 年及以上	23.1		201—500 人	13.8
				500 人以上	33.8

3.3 团队内部评价一致性检验

本文研究对象为团队层面,为了调研数据更能反映团队真实情况,本次调研中团队价值观异质性、知识共享和团队绩效的测量均以个体为单位进行,因此需要将个体数据加总,转化成团队数据。本文采用 rwg 指数对团队内部评价一致性进行检验。经检验,价值观异质性、知识共享和团队绩效三个变量的 rwg 指数检验结果分别为 0.91、0.95 和 0.98,均大于 0.7,因此本次收集的数据,团队内部评价一致性较高,适合进行团队加总。

4 数据分析与假设检验

4.1 量表信度、效度检验

本文利用 SPSS19.0 对价值观异质量表、知识共享量表和团队绩效量表进行因素分析和信度检验。经检验,三个量表的 KMO 值分别为 0.770、0.805 和 0.891,均大于 0.7;各变量在共同因素上的因子载荷在 0.664 到 0.917 之间,均大于 0.5;三个量表可以解释的总变异量均大于 60%,因此各量表建构效度良好。此外,经检验三个量表的 Cronbach α 系数分别为 0.839、0.912 和 0.915,均大于 0.8,量表信度较好。

4.2 控制变量方差分析

本文选取团队规模和团队成立时间作为控制变量,因为经文献研究发现,这两个变量潜在影响成员间的沟通与合作方式,可能会对成员间的知识共享、团队绩效等研究的因变量产生潜在影响。利用单因素方差分析方法对上述两个控制变量进行分析,检验结果表明,团队规模和团队成立时间对知识共享和团队绩效的三个维度的 F 值均未达到显著水平。究其原因,本文调研样本分布相对集中,未能做到严格的分层抽样,因此控制变量的对本次研究的影响并不显著。

4.3 假设检验

建立团队异质性与团队绩效间的初始结构方程模型,经检验,除价值观异质性对任务绩效的路径系数显著性为 0.301 未达到 0.05 的显著水平外,其他路径系数均达到了 0.05 的显著水平,删除不显著路径后,根据 MI 修正指标对模型进行修正。修正后的标准化估计模型如图 4-1 所示,模型的适配度指标中 P 值 = 0.066 > 0.05, RMSEA 值 = 0.041 < 0.08, GFI 值 = 0.901 > 0.9, NFI、RFI、IFI、NNFI、CFI 等增值适配度指标均大于 0.9,因此模型的适配度良好。修正后的模型中,各条路径的标准化系数均达到了 0.05 的显著水平。因此,除价值观异质性对任务绩效的负向作用未得到支持外,假设 H1 的其他假设均得到支持,即 H1a 成立、H1b 成立、H1c 部分成立。

加入中介变量知识共享,建立团队异质性、知识共享与团队绩效间的初始饱和模型。经检验,社会分类异质性对创新绩效和成员满意度、信息异质性对创新绩效和成员满意度、价值观异质性对创新绩效五条路径的系数分别为 0.419、0.555、0.183、0.160 和 0.445,均未达到 0.05 的显著性水平。将上述五条不显著路径删除后,根据修正指标对模型进行修订。修正后的模型如图 4-2 所示,模型的适配度指标中 P 值 = 0.059 > 0.05, RMSEA 值 = 0.037 < 0.08, PGFI 值 = 0.648 > 0.5, 卡方自由度比 = 1.174 < 2.00, NFI、IFI、NNFI、CFI 等增值适配度指标均大于 0.9,因此模型的适配度良好。

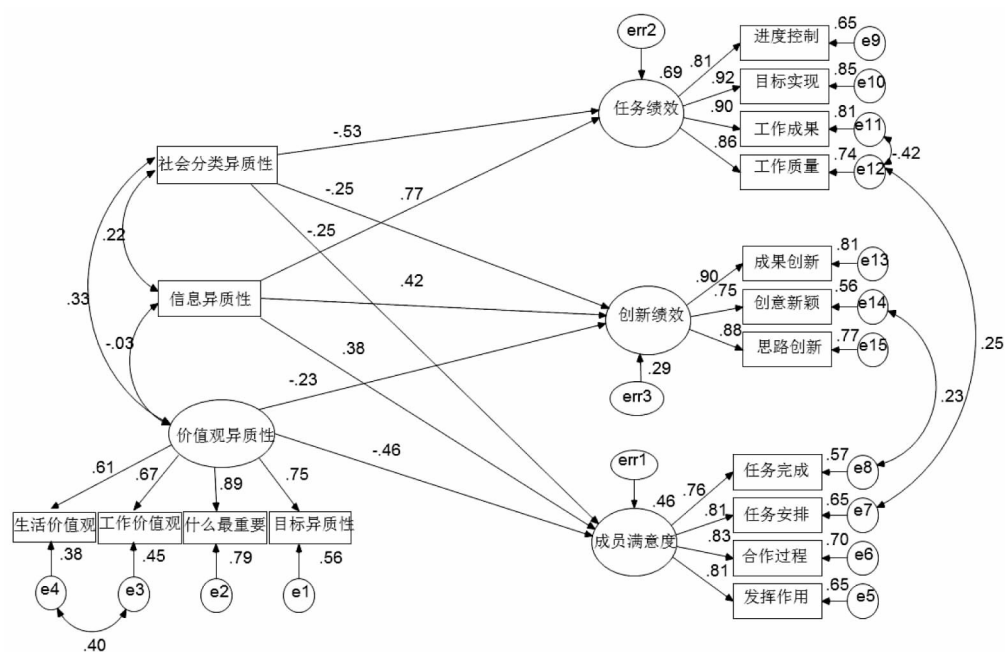


图 4-1 修订后团队异质性与团队绩效的标准化估计模型图

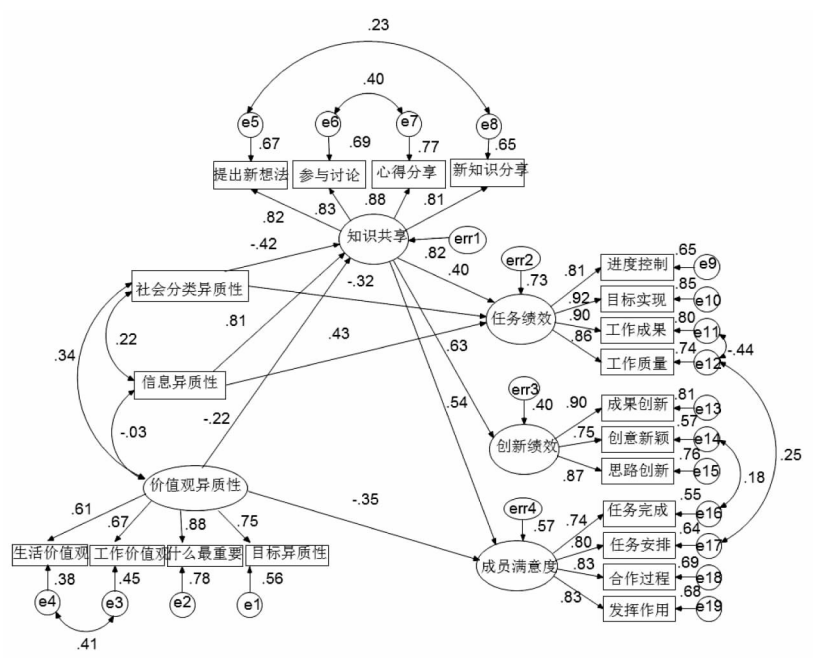


图 4-2 修订后团队异质性、知识共享与团队绩效的标准化估计模型图

修正后团队异质性、知识共享与团队绩效的标准化路径系数中社会分类异质性、信息异质性和价值观异质性对知识共享的标准化路径系数分别为 -0.425 、 0.813 和 -0.224 ，均达到了 0.05 的显著水平，且未出现与理论期望参数符号相反的路径，因此假设 H2a、H2b、H2c 均成立。同理，修正后的模型中，知识共享对任务绩效、创新绩效和成员满意度的标准化路径系数分别为 0.380 、 0.984 和 0.690 ，均达到了 0.05 的显著水平，因此假设 H3a、H3b、H3c 也得到支持。

本文采用 Baron 和 Kenny(1986)提出的中介变量判断标准对知识共享的中介效应进行检验。由表 4-1 可知,加入知识共享变量后社会分类异质性对任务绩效、信息异质性对任务绩效、价值观异质性对成员满意度的路径系数均有所降低,因此知识共享在这三条路径中起到部分中介作用;同理,未加入知识共享变量前团队异质性(社会分类异质性、信息异质性和价值观异质性)对创新绩效的三条路径、社会分类异质性和信息异质性对成员满意度的两条路径均达显著水平,而加入知识共享变量后,以上五条路径系数均变为不显著,因此知识共享在这五条路径中起到了完全中介作用。综上,假设 H4 成立。

表 4-1 知识共享的中介效应汇总

团队异质性与团队绩效关系	初始路径系数	加入知识共享后路径系数	中介效应
社会分类异质性→任务绩效	-0.527	-0.323	部分中介效应
信息异质性→任务绩效	0.767	0.429	部分中介效应
价值观异质性→任务绩效	不显著		
社会分类异质性→创新绩效	-0.253	不显著	完全中介效应
信息异质性→创新绩效	0.419	不显著	完全中介效应
价值观异质性→创新绩效	-0.227	不显著	完全中介效应
社会分类异质性→成员满意度	-0.251	不显著	完全中介效应
信息异质性→成员满意度	0.38	不显著	完全中介效应
价值观异质性→成员满意度	-0.458	-0.353	部分中介效应

5 结论与建议

本文研究发现,社会分类异质性对团队绩效和知识共享均起到了负向作用;信息异质性对团队绩效和知识共享均起到正向作用;价值观异质性对知识共享和团队绩效中的创新绩效和成员满意度均起到负向作用,而对任务绩效的负向作用未达到显著水平;知识共享在团队异质性与团队绩效间起到了中介作用。

根据实证研究结论,本文为如何提高文化创意团队绩效提出如下建议:

第一,合理配备文创团队成员。首先,在文化创意团队组建过程中,要根据团队目标,挑选合适的团队成员,保证团队成员在专长、教育背景、职能部门经历等方面的异质性。其次,为了降低成员价值观异质性对团队绩效的阻碍作用,建议利用价值观测试量表挑选具有相同价值观的团队成员。此外,为了最小化社会分类异质性对团队绩效的负向作用,建议团队领导者增加团队建设活动的频次,促进不同社会分类组别成员间的信任与非正式沟通。

第二,促进文创团队知识共享行为。首先,建议文创团队构建合作、共享的团队文化。其次,建议文创团队构建畅通的团队沟通机制。最后,建议文创团队构建完善的共享激励机制,增强成员的知识共享意愿,提升团队总体绩效。

参 考 文 献

- [1] Van Knippenberg D, De Dreu C W, De Dreu A C. Work Group Diversity and Group Performance: An Integrative Model and Research Agenda [J]. *Journal of Applied Psychology*, 2004, 89 (6): 1008—1022
- [2] Jehn K A, Northcraft G B, Neale M A. Why Differences Make a Difference: A Field Study of Diversity, Conflict, and Performance in Workgroups[J]. *Administrative Science Quarterly*, 1999, 44(4): 741—763
- [3] Roberge M, Van Dick R. Recognizing the benefits of diversity: When and how does diversity increase group performance? [J]. *Human Resource Management Review*, 2010, (20): 295—308
- [4] 邹子珩. 文化创意企业的创新成长过程研究[D]. 陕西:西北大学,2011
- [5] 张大为. 员工关系对知识共享的影响研究[D]. 大连:大连理工大学,2011
- [6] 李夏楠,倪旭东. 基于团队知识异质性结构的知识整合研究[J]. *科技进步与对策*,2012,29(11):132—137
- [7] 盛晓伟. 知识共享与创新绩效的实证研究——基于创新型企业的数据[D]. 成都:西南财经大学,2012
- [8] 刘天睿. 北京市文化创意产业高端人才培养模式研究[D]. 北京:中央民族大学,2012

Research on the Relationships among Team Heterogeneity, Knowledge Sharing and Team Performance in Creative Enterprises

Qin Shuang

(School of Economics and Management, North China Univ. of Tech. , Beijing, 100144)

Abstract According to the characteristics of creative team, this paper established a theoretical model of team diversity, knowledge sharing and team performance by introducing knowledge sharing as the medium variable. Hypotheses were tested through a sample of 130 creative teams, using the software of SPSS19.0 and AMOS17.0 (structural equation model software). Results showed that social category heterogeneity has a significant negative effect on team performance and knowledge sharing; information heterogeneity has a significant positive effect on team performance and knowledge sharing; value heterogeneity has a significant negative effect on knowledge sharing, team innovation performance and member satisfaction, but it did not have a significant negative effect on task performance; knowledge sharing played a mediating role between team heterogeneity and team performance. Finally, some suggestions for creative team management were put forward, according to the results of the study.

Key words creative enterprises; team heterogeneity; knowledge sharing; team performance

基于社会网络分析的建筑工业化利益相关者关系研究

研究生 李晓桐

导师 张铁山 纪颖波

(北方工业大学经济管理学院 北京 100144)

摘 要 本文考虑我国建筑工业化发展过程中参与方众多和协同创新困难等问题,通过深度访谈和问卷调查,收集多个工业化住宅项目利益相关者的关系数据,利用社会网络分析深入研究利益相关者关系网络的结构特征,并探讨与技术、管理、政策等影响因素之间的交互作用,构建建筑工业化利益相关者关系网络模型,揭示建筑工业化利益相关者关系网络的运行规律和影响因素,认为具备研发、设计、施工、运营一体化能力的建筑工业化企业集团在网络中具有核心地位,政府出台的激励政策也直接决定着建筑工业化发展进程。并为进一步提升中国建筑工业化协同创新绩效和能力提出建议。

关键词 建筑工业化;利益相关者;社会网络分析;影响因素

作者简介 李晓桐,女,1989年出生,硕士研究生,主要研究方向为技术创新与价值工程。

E-mail: lixiaotong75@163.com

1 绪论

改革开放以来,我国的建筑产业持续快速发展,产业规模不断壮大,建筑业的生产总值在2003—2014年间由近7500亿元上升到176713.4亿元,年增长率30.12%,形成了涵盖建材、房地产、工程建设、运营维护等产业链条,成为国民经济支柱产业。同时,我国建筑业总体上发展模式粗放、科技含量不足、资源耗费量大、劳动力生产率较低等诸多问题,严重制约建筑业可持续健康发展。

建筑工业化一般是指以设计标准化、生产工厂化、施工装配化、管理信息化为主要特征,形成完整、有机的产业链,实现建筑产品节能、环保、全生命周期价值最大化的可持续发展的新型建筑生产方式,是转变中国建筑业发展方式的重要战略,对这一领域的科学研究和工程实践成为行业热点问题。

建筑工业化的发展获得了国家及各省市政府部门的高度关注和支持。近年来,相继颁布鼓励建筑工业化发展的优惠政策,在保障性住房、部分公共建筑建设中强制采用工业化建造方式。但是目前我国建筑工业化仍处于起步阶段,存在诸多现实瓶颈问题,如技术成果转化率、组织信息传递不畅,协同创新能力弱、推进效果不佳等。

建筑工业化的发展涉及政府、建设单位、设计单位、构件生产单位、施工总承包单位、科研院校等多个利益相关者,涉及不同组织间不同专业技术的协调与集成应用。利益相关者之间的关系是一种客观存在的社会结构,任何类型的社会关系均会对其他利益相关者的行为产生积极或消极的影响,这就构成了一个复杂的社会网络系统。

社会网络分析是作为研究组织间相互作用的有利工具,能够探究隐藏在复杂社会系统之下深层网络的作用规律。在建筑工业化发展进程中,社会网络分析能够有效地表达关系网络的形态、属性特征以及采取行为策略时的相互影响。分析利益相关者的社会网络系统,对建筑工业化的发展具有重要的

促进作用,是当前建设管理领域的前沿问题。

2 建筑工业化利益相关者关系网络模型的构建

2.1 关系网络模型要素的界定

建筑类企业和相关机构(科研机构、中介机构、政府等)在特定的政策环境下,通过彼此间的沟通和交流,建立起来的相对稳定的、能够促进建筑工业化顺利进行的、正式或非正式联系的总和,构成了建筑工业化利益相关者关系网络。利益相关者及其社会关系是关系网络的核心要素。

(1)利益相关者的确定

在社会网络中,各个节点代表行动者,即利益相关者。滚雪球式可以根据某些研究特征对样本数量进行筛选,适用于在大量空间内寻找一些较少的分析样本,能够为构建关系网络图时明确利益相关者提供有效的手段。利益相关者的识别从时间维、任务维、过程维三个角度进行,即影响或受影响的组织或个人,主要包括:建设单位、设计单位、总承包单位、专业分包单位、构件生产单位、监理单位、供应商、政府、科研院校、中介机构和金融机构。

(2)利益相关者之间的关系确定

在社会网络中,各连接线代表利益相关者之间的关联关系。根据利益相关者关联的内容不同,可以表示为:需求提出、资源提供、信息提供、方法提供等不同形式。关系有正式和非正式之分,也有强弱之分。因此,本文从交流强度、依赖度、承诺、信任四个维度来测度利益相关者之间的关系强度,构建一个赋权网络。

2.2 网络特征测度

从整体网络、个体网络两个角度对构建的关系网络进行分析。整体网络分析从网络全局的视角研究关系网络,探讨建筑工业化利益相关者关系网络结构特征,考察技术、知识、信息、资金等资源是否在网络中充分高效地流动,寻求适合建筑工业化发展的整体网络结构,主要选取密度(Density)、中心势(Centralization)、凝聚子群密度(External-Internal Index)、块模型(Block Models)指标。个体网络分析主要考察两个方面,一方面分析各利益相关者在关系网络中所处的位置、拥有的权利及对关系网络产生的影响力;另一方面分析各利益相关者对其他利益相关者的影响力和制约力,主要选取三个维度的中心度来测度个体网络,即点度中心度(Degree Centrality)、中间中心度(Betweenness Centrality)和接近中心度(Closeness Centrality)。

3 案例分析

建筑工业化的发展,大都随着住宅数量的需求、居民对住宅质量和功能要求的提高而发展起来的。“十二五”期间,我国政府提出 3600 万套保障房建设目标,因此,采用工业化方式建造保障性住房成为了推进建筑工业化发展的重要突破口。

3.1 案例基本情况

以多个采用工业化建造方式的保障性住房建设项目为研究案例,将其中每个利益相关者作为关系网络中的行动者,采用滚雪球式问卷调查法,询问“哪些利益相关者与本项目有关系”和“利益相关者之间的关系强度”,获取相关数据,构建邻接矩阵,得到建筑工业化利益相关者关系网络图。

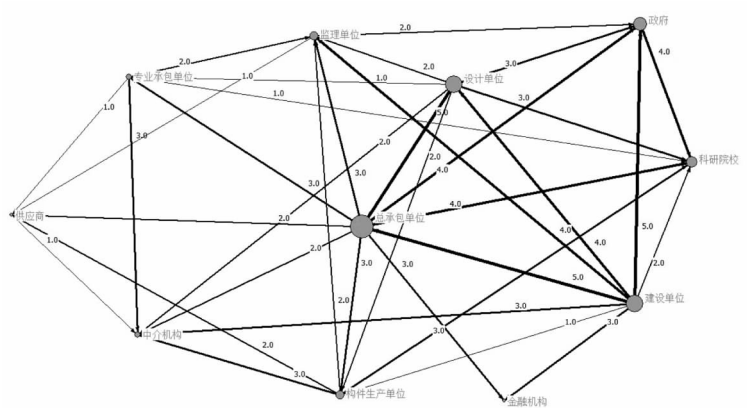


图 3-1 建筑工业化利益相关者关系网络图

利用 UCINET 软件的 MDS 功能对图做进一步处理。从图可以看出,总承包单位、建设单位、设计单位聚集到了关系网络的中心位置;而其他利益相关者则分散在关系网络的边缘位置,两级分化的核心—边缘结构非常明显。

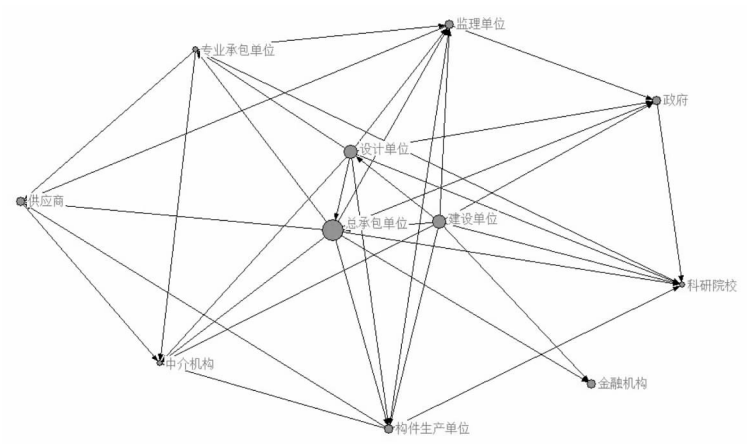


图 3-2 建筑工业化利益相关者关系网络核心—边缘 MDS 图

3.2 个体网络分析

(1)点度中心度

总承包单位(68.000%)、建设单位(54.000%)、设计单位(44.000%)具有较高的点度中心度,与较多的利益相关者保持着直接且较强的关系,占据网络的战略优势地位,在关系网络中具有较高的影响力。当前,以政府引导、企业参与的建筑工业化发展现状,使政府(36.000%)对其他利益相关者具有一定的影响力。而高等院校(34.000%)作为外部新知识、新技术的供给机构,能够推动知识、技术、信息等资源在关系网络中的扩散。

(2)中间中心度

中间中心度的值整体较低,大多数利益相关者并不具有控制资源的能力,整体网络不够稳定,缺少必要的资源,易出现结构洞。中间中心度最高的是总承包单位(2.583%),处于网络连接的核心地位,说明其掌握了较多的技术、知识、信息等资源,更能控制其他利益相关者。而供应商、高等院校、中介机构的中间中心度为 0,几乎不具备控制资源的能力。

(3)接近中心度

中介机构(24.390%,9.091%)、科研院校(23.560%,9.890%)具有较高点出度和较低点入度,表明其依赖于其他利益相关者提供资源,处于被动地位,但能以最短路径到达其他利益相关者,因此在知识、技术传递中处于核心的地位。总承包单位(9.091%,83.333%),建设单位(10.000%,45.455%)具有较低点出度和较高点入度,在资源输出上较少依赖其他利益相关者,是网络的中心节点。

3.3 整体网络分析

(1)密度和中心势

建筑工业化利益相关者关系网络的密度为0.2767,即27.67%的可能关系存在。外向点度中心势(31.313%)明显小于内向点度中心势(45.250%),说明利用网络资源的利益相关者比输出资源的利益相关者权力分布更加不均衡。中间中心势数值表明,网络中存在能够有效控制资源的利益相关者(19.400%)。整个关系网络集聚程度不高,但权力和资源在网络中比较集中。

(2)凝聚子群密度

建筑工业化利益相关者关系网络的凝聚子群密度较显著,达到了0.743,表明该网络中存在较严重的小团体现象。凝聚子群的规模较小,利益相关者的合作范围并不是很广,大部分利益相关者选择与其有长期合作经历的利益相关者进行合作,从一定程度上来说,这种情况影响技术、知识等资源在网络中的传递,不利于建筑工业化的发展。

(3)块模型

该关系网络被划分为2个团体。团体1是由总承包单位、建设单位、设计单位、政府以及科研院校所组成;团体2是由专业分包单位、构件生产单位、监理单位、供应商、中介机构以及金融机构所组成,呈现内紧外松的状态,即处于凝聚子群内部的利益相关者之间联系较频密,而处于子群外部的利益相关者则不能得到足够的信息和合作机会。中心度较高的利益相关者之间联结的战略同盟,如构建政产学研协同创新平台,比普通组织中任意联结的同盟对网络本身更具有影响力。

4 建筑工业化利益相关者关系的影响因素及假设检验

建筑工业化发展中存在着一定的经济、观念、政策等方面的制约。从政策、技术、管理因素探讨其对建筑工业化利益相关者关系的影响。通过文献研究和实证研究法,提取了13个可能影响建筑工业化利益相关者关系的影响因素,并提出假设。

表 4—1 建筑工业化利益相关者关系的影响因素

一级指标	二级指标	三级指标
政策因素	发展规划	保障性住房建设(I1)
	鼓励政策	土地政策(I2),金融政策(I3),税收政策(I4),其他政策(I5)
	标准规范	标准文件(I6)
技术因素	技术研发	人员、资金支持及时性(I7),学习吸收能力(I8)
	技术复杂性	专业协同(I9),技术集成度(I10)
管理因素	管理创新	BIM技术的应用(I11),集成化管理(I12)
	商业模式	全产业链发展模式(I13)

4.1 研究假设

(1)政策因素

建筑工业化的发展与政府政策的扶持和引导密切相关。从我国各省市发展情况来看,我国建筑工业化的推进工作主要以技术标准支持、政府政策鼓励为主,同时引导企业,开展建筑工业化示范试点,并在保障性住房建设中强制采用工业化建造方式。

在建筑工业化发展初期,政府制定了土地政策(面积奖励)、金融政策(信贷支持)、税收政策(税收优惠)、其他政策(减免基础设施配套费、建筑节能专项扶持资金、优先评优、优先中标)等,同时建立健全项目建设管理制度(技术标准、评价标准、标准图集、清单定额等),规范建筑工业化发展。综上,提出研究假设:各政策因素对网络密度显著正相关。

(2)技术因素

建筑工业化的发展不仅需要政府政策的引导,同时也需要利益相关者之间协同创新。技术是推进建筑工业化的基础,但工业化建造技术的研发需要大量的投资和人才,而且研发成果具有很强的外部性,导致企业对技术集成创新缺乏动力。引进和吸收国内外相近技术,是初期企业发展建筑工业化的重要途径,因此企业的学习吸收能力尤为重要。工业化建造方式不同于传统建造方式的最大特点是体现全生命周期的理念,将设计施工一体化,强调不同专业之间的协同。综上,提出研究假设:各技术因素对点度中心度显著正相关。

(3)管理因素

管理是建筑工业化推行的关键。将 BIM 技术与工业化深度融合,能促进建筑工业化的长足发展。建筑工业化具有多专业集成、全生命周期集成、项目参与方集成和管理目标集成等特点,集成化管理能够解决信息传递过程的漏斗效应和信息孤岛现象。现在是建筑业和制造业大融合的阶段,企业需向全产业链模式发展,以设计为龙头,整合产业链上优势资源,方能实现企业转型升级。综上,提出研究假设:各管理因素对点度中心度显著正相关。

4.2 假设检验

以上假设在检验时主要分为以下两类:一类是政策因素对建筑工业化利益相关者关系网络的影响,即点—关系混合(Mixed Dyadic/Nodal)层次的假设检验,采用关系列联表分析(Relational Contingency-Table)方法;另一类是技术因素、管理因素对利益相关者属性的影响,即点层次属性(Node-level)数据的假设检验,采用回归分析(Regression)方法。假设结果表明,大部分假设关系通过了验证。

表 4—2 研究假设检验结果

假设类型	假设	假设路径	检验结果
政策因素 与网络密度	H1	保障性住房建设对网络密度显著正相关	成立
	H2	土地政策对网络密度显著正相关	成立
	H3	金融政策对网络密度显著正相关	不成立
	H4	税收政策对网络密度显著正相关	成立
	H5	其他政策对网络密度显著正相关	成立
	H6	标准文件对网络密度显著正相关	成立

假设类型	假设	假设路径	检验结果
技术因素 与网络位置	H7	人员、资金支持及时性对点度中心度显著正相关	成立
	H8	学习吸收能力对点度中心度显著正相关	成立
	H9	专业协同对点度中心度显著正相关	成立
	H10	技术集成度对点度中心度显著正相关	成立
管理因素 与网络位置	H11	BIM 技术的应用对点度中心度显著正相关	成立
	H12	集成化管理对点度中心度显著正相关	成立
	H13	全产业链发展模式对点度中心度显著正相关	成立

5 结论和研究展望

借鉴已有研究成果,结合我国建筑工业化的现实特点,运用社会网络分析刻画了建筑工业化利益相关者关系网络,进而再研究政策、技术、管理因素对网络结构和网络关系的影响。

从整体网络和个体网络两方面,对建筑工业化利益相关者关系网络进行分析。关系网络凝聚力较弱,小团体现象较严重,权力和资源较集中;总承包单位、建设单位、设计单位、政府是关系网络中的重要利益相关者,在网络中具有一定影响力。因此,应通过加强与外部合作,为整体网络吸收新的利益相关者,推进网络由疏到密的演化进程。政府层面,加强政策扶持,完善相应管理制度,推进建筑工业化稳序发展;企业层面,加强工业化技术、管理创新活动,构建产学研协同创新平台,形成集研发、设计、施工、运营等于一体的建筑工业化企业集团,其在网络的核心地位,也将会对其他利益相关者、关系网络演化产生重要影响。

本文构建了一个建筑工业化利益相关者关系网络模型,但该建模和分析方法在建筑工业化领域尚处于起步阶段,其中的更多细节问题亟待展开,例如需要进一步考虑网络的动态演变及发展模式,并扩展其他类型的建筑工业化项目进行分析。

参 考 文 献

[1] 纪颖波. 建筑工业化发展研究[M]. 中国建筑工业出版社, 2011

[2] Girmscheid G, Rinas T. Business Design Modeling for Industrialization in Construction: Cooperative Approach [J]. Journal of Architectural Engineering, 2012, 18(2): 164—175

[3] 纪颖波, 李晓桐. 建筑工业化发展的政策建议[J]. 施工企业管理, 2014, (5): 60—61

[4] XU Q, SU Z, WANG X. Factors Analysis of Supporting Environment for Construction Industrialization Based on System Dynamics[J]. Journal of Engineering Management, 2012, 4: 007

[5] 丁荣贵. 项目利益相关方及其需求的识别[J]. 项目管理技术, 2008, (1): 73—76

[6] Jaillon L, Poon C S. The evolution of prefabricated residential building systems in Hong Kong: A review of the public and the private sector[J]. Automation in Construction, 2009, 18(3): 239—248

[7] Rutten M E J, Dorée A G, Halman J I M. Innovation and interorganizational cooperation: a synthesis of literature [J]. Construction Innovation, 2009, 9(3): 285—297

[8] Salas J, Blázquez A, Oteiza I, et al. The technical approvals as potential incentive of construction Industrialisation [J]. Informes de la Construcción, 2013, 65(531): 275—288

- [9] Li H, Guo H L, Skitmore M, et al. Rethinking prefabricated construction management using the VP - based IKEA model in Hong Kong[J]. Construction Management and Economics, 2011, 29(3): 233-245
- [10] 刘军. 整体网分析讲义[M]. 上海: 格致出版社, 上海人民出版社, 2009

Social Network Analysis of Stakeholder Relationships during Construction Industrialization in China

Li Xiaotong

(School of Economics and Management, North China Univ. of Tech., Beijing, 100144)

Abstract Modular and off-site construction has garnered world-wide attention, and particularly the construction industrialization is about to launch in China, which will be a transition for stakeholders within the construction industry. Furthermore, construction industrialization will alter relationships among stakeholders to a certain degree. In this research, the factors influencing construction industrialization in China are analyzed from the perspectives of technology, management, and policy; based on the analysis, stakeholders are identified and relationships are mapped through social network analysis. Furthermore, stakeholder relationships are evaluated and analyzed based on the data, which are collected through interviews and questionnaires administered to stakeholders of industrialization projects. This research aims to investigate the underlying mechanism of stakeholder relationships during construction industrialization in China, and to provide recommendations for promoting the performance and capacity of construction industrialization in China.

Key words construction industrialization; stakeholders; social network analysis; influencing factors

基于生态学的京津冀地区创新适宜度评价研究

研究生 傅建敏

导师 张铁山 张淑谦

(北方工业大学经济管理学院 北京 100144)

摘 要 京津冀是环渤海区域中最具影响力的地区,其发展直接关乎未来华北区域的整体竞争实力。在京津冀一体化建设背景下,实现京津冀地区创新资源优势互补、提升区域创新水平问题具有重要的战略意义。本文在国内外研究成果基础上,结合京津冀区域创新生态系统发展实际,引入生态适宜度概念,构建了区域创新生态系统创新适宜度评价指标体系,并对京津冀地区的创新适宜度进行评价分析,并在此基础上提出对策建议,以期为促进京津冀优势互补和一体化发展提供科学支持和决策依据。

关键词 区域创新生态系统;创新生态位;创新适宜度;评价分析

作者简介 傅建敏,女,1988 年出生,硕士研究生,主要研究方向为技术创新与价值工程。

E-mail: fujm2015@163.com

1 绪论

十八大三中全会以来,京津冀一体化发展已上升到国家战略层面。京津冀是环渤海区域中最具影响力的地区,其发展直接关乎未来华北区域的整体竞争实力。要实现京津冀一体化发展就必须坚持优势互补、科学持续的协同发展。目前对区域创新生态系统的研究大都针对创新生态系统本身,而对创新环境对创新活动的影响作用的研究相对不足。生态学作为一门研究生物与自然环境关系的学科,其理论方法已被广泛应用到不同领域,而且将生态学的理论方法应用于经济学研究具有一定的科学性和合理性。而创新适宜度是指创新活动所需资源的现实生态位与最适生态位的匹配程度;创新适宜度进化动量反映的是创新适宜度的提升空间。

因此,本文通过引入生态位的理论与方法,构建了区域创新适宜度评价指标体系,并结合生态位适宜度评估模型,对京津冀地区创新适宜度及其进化动量进行测度,明确了各地区在不同创新生态因子上的优劣势及实行对策后的效果情况,并针对京津冀地区创新发展中存在的问题提出对策建议,这为实现京津冀优势互补和一体化发展政策制定提供了科学依据。

2 区域创新生态系统相关基础理论及文献综述

2.1 生态位适宜度相关理论及研究现状

关于生态位(Ecological Niche)的研究,最具代表性的是 Hutchinson 提出的“N 维超体积生态位”,Hutchinson 将生态位定义为它是生物单元在生态环境中的位置和功能地位的集合体,并用数学中点集思想将现实生态环境抽象为一个多维的空间,物种的不同生存影响因素即为该空间的不同纬度。那么在这一定义中既体现了物种生态位的空间性又体现了其功能性,其数学化的界定方式为之后的生态位定量研究奠定了方法基础。此后我国学者李自珍将生态位描述为“是生物种的生态学特征的

定量描述,其内涵包含了种对环境的需求、环境对种的影响以及种与所在群落其他物种之间的关系”^[1]。这阐释为生态位的定量研究提供了明确的研究思路,为生态位理论在其他学科中的拓展应用奠定了关键性的理论基础。

关于生态位适宜度,最早由李自珍等人提出这一概念,他将生物种的生态位适宜度定义为“一个种居住地的现实生境条件与最适生境条件之间的贴近程度^[1]”,并在深入剖析生态位内涵的基础上,提出了测度生态位适宜度的评估模型,实现了对生态位适宜度的量化研究。目前该生态位适宜度评估模型已广泛应用于农业种植、土地资源、城市系统等研究领域。

2.2 区域创新系统的研究现状及趋势

经过大量的文献研究发现,创新生态系统理论是一个比较新的理论,是管理学与生态学结合的交叉性学科。对创新生态系统的研究与应用是对产业集群技术创新研究领域的新拓展,将对区域创新效率提高以及社会资源的有效利用等方面的研究提供新的思路和方法。但目前国内外研究主要集中于其概念界定、运行机制调节机制等管理机制研究等方面,并未形成比较完整的理论体系,一方面将创新生态理论与实际问题结合的研究比较少见;另一方面,已有的评价研究只考虑创新系统组成要素对创新能力的影响作用,而忽略了组成要素之间的相互关系对区域创新系统的作用,因此全面深入分析如何提高区域创新能力方面研究不足。因此这也是今后区域创新系统评价研究的一个热点问题。

3 区域创新生态系统结构模型的构建

3.1 区域创新系统与自然生态系统的对比分析

目前生态学的理论方法已被广泛应用到不同领域,从已有研究和实践总结可见,生态学跨学科应用的有效条件有两种:“一是研究对象与生态环境有密切关系,他们相互影响、相互制约;二是研究对象具有生态系统的特征”^[4]。对于区域创新系统的生态学特性,黄鲁成、傅羿芳等人从构成要素和系统内行为两个角度对区域创新系统与生态系统进行了较全面的对比分析。综合以上研究,本研究提出以下对比分析(表 3-1)。由此可知,区域创新系统研究适合应用生态学理论与方法。

表 3-1 区域创新系统与生态系统的各要素对比分析

生态系统要素	内涵	区域创新系统要素	内涵
生物个体	完整的具有生长、发育和繁殖等功能的生物有机体	创新组织	独立的个体创新单位。如企业、高校、科研单位等
物种	在生物圈内,具有相同基因频率,形态和生理特征的生物个体的集合	创新复合组织	创新主体与创新相关体
种群	在一定时空内,同一物种的个体的集合体	创新种群	一定地域内,具有相同资源和能力、技术及同类产品的创新实体的集合
群落	不同生物种群的集合	创新群落	不同创新种群的综合体
生态系统	在一定时空内,生物群落与环境之间不断进行物质、能量和信息交换而形成的统一体	创新生态系统	在一定地域内,创新综合体与环境相互作用而形成的具有内在创新能力与外在创新条件、协调互动的统一体

生态系统要素	内涵	区域创新系统要素	内涵
生产者	用无机物制造有机物的生物	创新主体	利用各种资源进行技术创新的企业和科研机构
消费者	消费生产者制造的有机物者	创新成果消费者	吸收或使用创新成果者
食物网	物种的营养位置及关系	创新网	各创新种群通过生产、使用创新成果而形成的创新技术传递与对接关系
生态位	在特定时空内,一个生物单位(如种群)对各类资源的利用和对环境适应性的总和	创新生态位	在特定区域内,创新组织对各类资源的利用和对环境适应性的总和
生态因子	对生物生长有影响的环境要素	创新生态因子	对各创新组织有影响的环境要素
生境	生物个体和群体生活的具体生态环境	创新生态环境	创新主体所处的具体创新生态环境,如区域人文环境、基础设施、政策环境等
能量流动	能量在生态系统中的流转	能量流动	能量在创新系统中的流动
物质流动	物质在生态系统中的流转	物质流动	热能和动力在创新系统中的流转
信息流动	信息在生态系统中的流转	知识流动	创新知识的产生、扩散
进化	种群发展变化以满足新环境的机制	渐进性创新	对现有技术的渐进、连续创新
突变	物种遗传物质的可遗传性改变	根本性创新	技术上的重大突破
生存	在争夺相同资源中存活	活力	适应变化的能力
适应	随自然环境变化而变化	应变	对变化的自然、社会环境做出反应
协同共进	为适应环境,各种群通过相互作用相互适应而共同进化	协同创新	在外部环境或内部因素的作用下,各创新组织既竞争又合作,协同发展

3.2 区域创新系统的生态学特征

学者们对区域创新系统的特征进行了研究并提出各自观点。顾新认为区域创新系统具有构成要素及结构的客观性、各具地域特色的多样性、整体性、自组织性和开放性等特点;而黄鲁成、张淑谦等则从动态角度认为区域创新系统具有适应性、协同性、相关性、有序性和再生性等特点。尽管不同学者对区域创新系统特征的分类不同,但其阐述的内容是相通的,本研究在综合已有研究成果基础上,认为区域创新系统具有与生态系统类似的一般特征:整体性、层次性、自适性和耗散性。

3.3 区域创新生态系统的结构分析及模型构建

基于对区域创新生态系统概念内涵及构成要素等的分析,本研究提出区域创新生态系统的空间结构模型如图 3—1 所示:

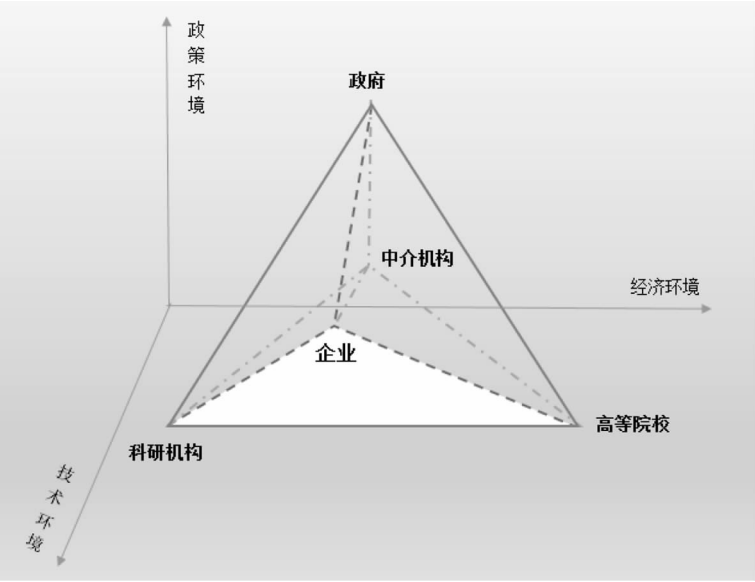


图 3-1 区域创新生态系统的结构模型

从以上结构模型可见,在一定区域内,由企业、科研机构和高等院校构成的创新主体,在政府及中介机构等创新支撑服务机构的支持下,从无边界的经济环境、技术环境及政策环境中获取创新资源以进行创新活动,同时在创新活动过程中或通过创新成果的产出流通和应用来促进所处创新环境的发展和优化^[5]。

4 创新适宜度评价指标体系及评价模型的构建

4.1 创新适宜度的概念及内涵

在现代生态学理论中,生态位是指“在特定时空内,一个生物单元(如种群)对各类资源的利用和对环境适应性的总和”^[1]。因此引用生态位的这一理论,并基于区域创新生态系统的生态学特性以及创新主体开展创新活动的实际过程,本文中,区域创新生态系统创新生态位是指由创新生态系统在经济环境、政策环境、技术环境以及创新主体自身创新能力等创新生态因子上的适宜区所形成的多维立体空间。

适宜度是指研究对象(物种单元)在一定区域或空间里生存或发展的适宜程度,其量化结果即可直观反映所研究对象在当前区域或空间中生存发展的适宜程度。在综合和吸收已有研究成果基础上,本研究认为区域创新生态系统创新适宜度(以下简称“创新适宜度”)是指在一定区域内,区域创新生态系统中的创新主体,在其所处的区域内开展创新活动时,创新环境所能提供的现实创新资源位(即现实创新生态位)与创新主体所需的最适创新资源位(即最适创新生态位)之间的贴近程度。

4.2 创新适宜度评价指标体系的构建原则

一个好的评价指标体系应是能够较全面的反映出所评价内容的指标的有序集合。因此本文在建立评价指标体系时,遵循了以下原则:(1)系统性原则;(2)客观性原则;(3)科学性与可行性相结合原则;(4)简明性原则。对指标体系的结构框架进行合理规划并甄选具体的测度指标,进而依据所构建的指标体系选定合适的评价方法以及数据来源。

4.3 构建创新适宜度评价指标体系

本研究在深入剖析区域创新生态系统生态学特性及结构、理解区域创新生态系统生态位适宜度内涵基础上,综合国内外相关评价指标体系文献,结合京津冀地区区域创新生态系统的发展实际,构建了一个包含创新主体、创新环境、创新主体与创新环境之间联系等 3 个一级指标和 26 个二级指标的区域创新生态系统适宜度评价指标体系如表 4—1 所示:

表 4—1 区域创新生态系统创新适宜度评价指标体系

测度目标	测度要素	测度指标	实测指标
区域 创新 生态 系统 创新 适宜 度	创新主体 B1	企业	规模以上工业企业数量 C11
			规模以上工业企业科研活动人员
			规模以上工业企业 R&D 经费内部
			规模以上工业企业新产品销售收
		科研机构	科研机构数量 C15
			科研机构 R&D 人员数量 C16
			科研机构 R&D 经费内部支出 C17
			科研机构 R&D 科研仪器设备支出
		高等院校	高等院校数量 C19
			高等院校初级及以上职称教师数
			高等院校 R&D 人员数量 C111
	创新环境 B2	政策环境	科学技术地方财政支出 C21
			国家财政性教育经费 C22
		经济环境	地区人均生产总值 C23
			固定资产投资额 C24
			居民消费价格指数 C25
		技术环境	每十万人大专以上学历人员数
			互联网宽带接入端口数 C27
			人均拥有公共图书馆藏书量 C28
			科技进步环境指数 C29
	创新主体与创新环 境的联系质量 B3	创新成果产出	获得授权专利数量 C31
			发表科技论文数量 C32
		创新成果流通	输出技术交易额 C33
			吸纳技术交易额 C34
		创新成果应用	高技术产业化指数 C35
			科技促进经济社会发展 C36

(1)创新主体

根据对创新适宜度内涵分析可知,创新适宜度的大小受创新主体对创新环境的适应能力影响,这种适应能力包括对创新资源的识别、吸收和利用能力,因此对创新适宜度的评价需包含对创新主体的创新能力的测度。因此本研究从创新规模、创新投入及创新持续力三方面反映创新主体对创新环境的适应能力。

(2)创新环境

创新环境是对创新活动有直接或间接影响的环境要素的总和,为创新主体开展创新活动提供所需

的各种创新资源和条件,主要含政策环境、经济环境和技术环境。

(3)创新主体与创新环境的联系

由于区域创新生态系统是一个由创新主体与创新环境相互作用而形成的复杂网络结构。其中既包含创新主体、创新环境等“可见的”点,又包含两者之间“看不见的”联系——即两者间的物质、信息及能量的交流活动。因此在研究区域创新生态系统创新适宜度时,应考虑创新主体与创新环境间的联系质量,这种联系质量主要体现在创新成果的产出、流通及应用上。

4.4 创新适宜度评价模型的选择与修正

4.4.1 创新适宜度评价方法的选择

李自珍提出的生态位适宜度评估模型是建立在对生态位适宜度概念内涵深入剖析基础上,以生态位理论为基础,用数学抽象方法来系统分析种群的最佳生态位与现实生态位之间的贴近程度。因此该模型的评价特点与创新生态位适宜度的研究目标一致。基于以上考虑,本研究结合区域创新系统研究的自身特点,及研究客观性、科学性要求,对原有的生态位适宜度评估模型进行指标权重修正,并用修正后的评估模型研究对象的创新适宜度进行评价分析。

根据创新生态位适宜度的概念内涵,其主要考察的是现实生态位与最适生态位之间的差距 φ_{ij} ,在此基础上对指标体系中各指标的这一差异值进行加权计算,即可求得生态位适宜度,如公式(1)

$$S_i = \frac{1}{m} \sum_{j=1}^n \frac{\min\{\varphi_{ij}\} + \max\{\varphi_{ij}\}}{\varphi_{ij} + \max\{\varphi_{ij}\}}, (i=1, 2, \dots, m) \quad (1)$$

其中: S_i 即为所求的创新适宜度; $a(0 < a < 1)$ 为模型参数。由以上评价模型可知,创新适宜度介于0到1之间,适宜度值越高说明现实资源生态位与最适生态位之间的差距越小,该系统越适宜开展创新活动。

此外,该模型从未来发展空间的角度,对各评价指标改善后将对整体适宜度产生的影响效果(即适宜度的提升空间)进行了计算。模型以进化动量的概念来描述这一影响效果,其计算公式如下:

$$Q_i = \sqrt{\frac{\sum_{j=1}^n |y'_{ij} - y_{aj}|}{n}}, (i=1, 2, \dots, m) \quad (2)$$

4.4.2 创新适宜度评价方法的修正

在已有研究中,学者们对生态位适宜度评估模型中的生态因子权重值均采用主观均等化赋值方式,即令 $\gamma_i = 1/m (i=1, 2, \dots, m)$ 。显然这种做法不符合各创新生态位指标对适宜度的实际影响程度。对此,为使本研究结果更接近实际情况,本研究采用层次分析法来确定适宜度评估模型中各创新生态位指标的权重 γ_i 。将创新适宜度评价模型修正为:

$$S_i = \gamma_i \sum_{j=1}^n \frac{\min\{\varphi_{ij}\} + a \max\{\varphi_{ij}\}}{\varphi_{ij} + a \max\{\varphi_{ij}\}}, (i=1, 2, \dots, m) \quad (3)$$

为保证研究结果的准确性和有效性,本研究采用层次分析法(AHP)与专家打分法相结合的方法来确定适宜度评价指标的权重值。依据区域创新适宜度评价指标体系构建判断矩阵,并在论文指导老师帮助下,邀请11位在区域创新研究及创新生态系统研究方面科研成果较突出的教授及副教授作为打分专家,依据较成熟的1—9相对重要比例标度进行打分。经权向量计算及一致性检验,各判断矩阵均通过一致性检验($CR < 0.1$)如表4—1,得适宜度评价指标体系的各指标权重 $\gamma_i = (0.04, 0.07, 0.11, 0.06, 0.03, 0.06, 0.10, 0.07, 0.02, 0.03, 0.07, 0.04, 0.04, 0.03, 0.02, 0.01, 0.05, 0.02, 0.02, 0.05, 0.01, 0.01, 0.02, 0.01, 0.01, 0.01)$,如下表4—2、表4—3。

表 4—1 准则层 B 对目标层 A 的权向量计算及一致性检验

创新适宜度	创新主体	创新环境	联系质量	按行相乘	开N次方	权重Wi	AWi	AWi/Wi	CI=(λ - n)/(n-1)	CR=CI/RI
创新主体	1	3	7	21	2.76	0.65	1.99	3.06	0.03	0.06
创新环境	1/3	1	5	1.67	1.19	0.28	0.85	3.06		
联系质量	1/7	1/5	1	0.03	0.31	0.07	0.22	3.06		

表 4—2 指标层次结构模型总排序计算

	B1	B2	B3	总排序CR=Σ BiCi j
	0.65	0.28	0.07	
C11	0.06			0.04
C12	0.10			0.07
C13	0.16			0.11
C14	0.09			0.06
C15	0.05			0.03
C16	0.09			0.06
C17	0.16			0.10
C18	0.10			0.07
C19	0.03			0.02
C110	0.05			0.03
C111	0.11			0.07
C21		0.15		0.04
C22		0.14		0.04
C23		0.09		0.03
C24		0.09		0.02
C25		0.04		0.01
C26		0.18		0.05
C27		0.08		0.02
C28		0.07		0.02
C29		0.17		0.05
C31			0.12	0.01
C32			0.11	0.01
C33			0.29	0.02
C34			0.20	0.01
C35			0.16	0.01
C36			0.12	0.01

表 4—3 指标层次结构模型总排序一致性检验

CRi	0.07	0.08	0.08
RIi	1.52	1.46	1.26
CIi	0.10	0.12	0.10
	CR=ΣBiCRi 0.07	RI=ΣBiRIi 1.48	CI=ΣBiCIi 0.11
	CR=CI/RI 0.07		

5 京津冀地区创新适宜度评价与分析

5.1 我国京津冀地区创新适宜度评价

本文根据已构建的区域创新生态系统创新适宜度评价指标体系,基于《中国科技统计年鉴》、《中国统计年鉴》、《高技术产业统计年鉴》及科技部发布的《全国科技进步监测报告》等权威数据支持,采用李自珍提出的生态位适宜度评估模型方法,以京津冀三地区域创新生态系统为研究对象,对京津冀地区2007年至2012年的创新适宜度情况进行了测度研究。对原始数据进行标准化处理,取各指标的最大值为最优生态位值,标准化后的指标数据如表5-1所示(以2011年数据为例):

表 5-1 京津冀地区区域创新生态系统创新适宜度实测指标的标准化值

实测指标	地方	2011 标准化结果			最适生态位
		北京	天津	河北	
企业					
规模以上企业数量		0.32	0.43	1.00	1.00
规模以上工业企业科研活动人员数量		0.91	0.94	1.00	1.00
规模以上工业企业 R&D 经费内部支出		0.78	1.00	0.75	1.00
规模以上工业企业新产品销售收入		0.91	1.00	0.50	1.00
科研机构					
科研机构数量		1.00	0.16	0.20	1.00
科研机构 R&D 人员数量		1.00	0.08	0.07	1.00
科研机构 R&D 经费内部支出		1.00	0.06	0.05	1.00
科研机构 R&D 科研仪器设备支出		1.00	0.05	0.06	1.00
高等院校					
高等院校数量		0.78	0.49	1.00	1.00
高等院校初级及以上职称教师数量		0.98	0.47	1.00	1.00
高等院校 R&D 人员数量		1.00	0.28	0.29	1.00
政策环境					
科学技术地方财政支出		1.00	0.05	0.06	1.00
国家财政性教育经费		0.91	0.40	1.00	1.00
经济环境					
地区人均生产总值		0.96	1.00	0.40	1.00
固定资产投资额		0.76	1.00	0.82	1.00
居民消费价格指数		1.00	0.99	1.00	1.00
技术环境					
每十万人大专以上学历人员数		1.00	0.41	0.54	1.00
互联网宽带接入端口数		0.70	0.32	1.00	1.00
人均拥有公共图书馆藏书量		0.95	1.00	0.24	1.00

实测指标	地方	2011 标准化结果			最适生态位
		北京	天津	河北	
科技进步环境指数		1.00	0.84	0.56	1.00
创新成果产出					
获得授权专利数量		1.00	0.34	0.27	1.00
发表科技论文数量		1.00	0.15	0.09	1.00
创新成果流通					
输出技术交易额		1.00	0.09	0.01	1.00
吸纳技术交易额		1.00	0.26	0.10	1.00
创新成果流通					
高技术产业化指数		0.97	1.00	0.55	1.00
科技促进经济社会发展		0.96	1.00	0.82	1.00

由修正后的适宜度评价模型(公式 3)计算可得京津冀地区的区域创新生态系统创新适宜度值及进化动量,如下表:

表 5—5 京津冀三地区域创新适宜度值

	2007	2008	2009	2010	2011	2012
北京	0.8377	0.8722	0.9031	0.8028	0.8295	0.8060
天津	0.5045	0.5055	0.5225	0.5292	0.5433	0.5398
河北	0.5015	0.5041	0.5171	0.5176	0.5157	0.5282

表 5—6 京津冀三地区域创新适宜度进化动量值

	2007	2008	2009	2010	2011	2012
北京	0.2756	0.2706	0.2738	0.2904	0.2854	0.3357
天津	0.6988	0.6924	0.6934	0.6887	0.6842	0.6836
河北	0.7095	0.6991	0.6938	0.7016	0.6966	0.6903

5.2 京津冀地区创新适宜度的评价结果分析

通过对京津冀地区区域创新生态系统创新适宜度的测度研究发现,整体而言,北京、天津、河北的创新适宜度变化较为稳定,其中北京地区的创新适宜度最高(均在 0.8 以上,最高达 0.9)、河北地区创新适宜度最低(约为 0.51),天津地区创新适宜度略高于河北,创新适宜度值约在 0.53 左右。经过量化处理,京津冀三地在各指标上的优劣势更清晰直观(如表 5—1)。

从区域整体角度看,京津冀三地分别在不同创新资源位上占有最优生态位,即在创新资源上各有优势,具备实现资源互补的基础条件。这为促进三地创新资源整合,实现京津冀区域的协同一体化发展奠定了基础。通过评价分析易见,该区域中北京在创新主体的创新能力方面较天津河北两地优势明显,而天津河北两地创新资源尚未被充分利用、具备更大发展空间,因此创新主体特别是创新企业和科研机构,可根据各地区的资源优势情况及自身发展需求,选择更适宜自身开展创新活动的区域以实现

更好更快的成长和发展。同时政府作为区域创新中的重要一方,相关政策的制定需要以明确区域资源现状及分布为必要基础,以上对京津冀三地的创新适宜度评价及其在各资源位上的分布分析,将为此必要基础提供量化支持。

从提升空间角度,基于在不同创新生态位指标上的现实创新生态位值与最优生态位值间的贴近程度,各地区可以通过改变政策措施,或增加创新投入或优化创新人文环境等方式创造“变化”,进而实现对其创新劣势的改善和创新优势的发挥。研究发现,京津冀区域作为一个整体,相较而言天津河北创新适宜度提升的空间更大,因此通过加大对天津、河北两地的创新生态系统的创新影响因子的调整和改善,这能够更显著有效的提高京津冀地区整体的创新适宜度、促进区域协同创新发展。

从创新生态因子层面,本文对京津冀地区的创新资源位情况进行了对比分析,以期对资源整合、区域整体适宜度提高提供深层次、准确的支持。分析发现:(1)与北京相比,天津河北地区的创新适宜度较低,其创新主要依赖于外部技术的引进,自主创新能力相对不足;(2)北京地区的创新成果商业化程度不高;(3)京津冀三地的科技人才储备不均;(4)各地政府对科技和教育的支持力度和侧重点不同。

6 提升京津冀地区创新适宜度的对策与建议

基于以上对于京津冀地区创新适宜度的评价分析及对京津冀三地的区域创新生态系统内部创新生态位的对比分析,本研究从管理学角度对如何进一步提升京津冀地区的创新适宜度提出以下几点建议:(1)构建并完善京津冀三地间的科技人才流动机制;(2)搭建创新资源与创新成果共享平台;(3)增强区域创新生态系统建设的政府支持力度。

区域创新生态系统的持续创新能力不仅有赖于该区域的基础性研究能力和应用性研究能力,也离不开当地的技术开发或服务性研究能力。因此政府在加大对区域创新生态系统建设的支持力度的同时,应注重政府资金投入的多元化和投入比重均衡,避免因“木桶效应”所造成的资源浪费,从而实现区域创新生态系统的均衡发展。而为切实促进京津冀一体化发展,国家或三地政府部门应从统筹规划、综合发展角度出发,制定和实施一系列的完善政策措施以增强京津冀三地的区域创新生态系统间的物质、能量和信息交流,弱化系统边界的“隔离作用”,促进三地区域创新生态系统的融合,进而充分发挥京津冀一体化的规模效应、提升京津冀地区的整体创新水平和国际竞争力。

7 结论与展望

本研究基于对生态位理论和区域创新系统理论,并结合京津冀区域创新生态系统的发展实际,构建了新的区域创新适宜度评价指标体系,并在对传统的生态位适宜度评估模型进行权重修正的基础上,以修正后的创新生态位适宜度评估模型为创新适宜度的测度方法,对京津冀地区的创新适宜度情况进行了测度和评价分析。

研究发现:(1)京津冀三地区的创新适宜度并非一成不变的而是基本呈现动态的相对稳定状态,其中北京地区创新适宜度值最高、其次是天津、河北末之。三地区在创新资源方面各有不同的自身优势,具备实现资源整合、协同创新的基础条件;(2)从进化动量角度,河北的提升空间最大、天津次之、北京最少,即相关部门可依据各地区在不同创新生态因子上的资源优劣势情况,通过有针对性的采取构建完善京津冀科技人才流动机制、搭建创新资源与创新成果共享平台、增强政府支持力度和支持侧重点等合适的创新管理策略,天津、河北的创新适宜度将会得到很大程度的提高。

本研究的研究方法及结果不仅实现了对所研究对象的创新适宜度的量化,而且为区域创新生态系统的评价研究提供了新的评价指标体系和新的研究方法和思路,具有重要的理论价值。此外,本研究结果有助于区域创新管理的决策者全面掌握各地区的创新现状及优劣势条件,对促进京津冀区域资源整合、创新协同发展,具有较强的实践价值。

参 考 文 献

- [1] 李自珍,赵松岭.生态位适宜度理论及其在作物生长系统中的应用[J].兰州大学学报(自然科学版),1993,29(4): 219—224
- [2] Metcalfe S, Ramlogan R. Innovation systems and the competitive process in developing economies [J]. The quarterly review of economics and finance, 2008, 48(2): 433—446
- [3] 傅羿芳,朱斌.高科技产业集群持续创新生态体系研究[J].科学学研究,2004(22):128—135
- [4] 黄鲁成.区域技术创新生态系统的特征[J].中国科技论坛,2003(1):23—26
- [6] 孙冰,周大铭.基于核心企业视角的企业技术创新生态系统构建[J].商业经济与管理,2011(11):36—42
- [7] 苗红,黄鲁成.区域技术创新生态系统健康评价研究[J].科技进步与对策,2008,(8)
- [8] 周青,陈畴镛.中国区域技术创新生态系统适宜度的实证研究[J].科学学研究,2008,(S1)
- [9] 袁千里.基于生态位适宜度理论的区域创新系统评价研究[J].经济研究导刊,2012(13):170—171
- [10] 张淑谦,傅建敏.区域创新生态系统组成及其结构模型研究[J].新西部,2014(06):42

Innovation suitability evaluation and analysis of Beijing-Tianjin-Hebei region based on the ecology theory

Fu Jianmin

(School of Economics and Management, North China Univ. of Tech. , Beijing, 100144)

Abstract In recent years, scholars have conducted a lot of research on innovation capacity of innovation ecosystem, evaluation of system's health. But studies have generally targeted innovation ecosystem itself without attention to innovative environmental effects on innovative activity, and research has failed to solve the problem of how to improve the creative ability. Therefore, this article based on ecology niche theory, build a innovation suitability assessment index system of regional innovation ecosystem, with ecological suitability evaluation model for research, to measure and analysis the innovation suitability of Beijing-Tianjin-Hebei region. Research results will assist policy makers fully grasp innovation in the regions, specific advantages and disadvantages in the innovative activities in all regions, to develop and implement effective policy measures to enhance regional innovation capability to provide a scientific basis.

Key words regional innovation ecosystem; innovation niche; innovation suitability; evaluation and analysis

价值工程在食品冷链物流企业 低碳运营中的应用研究

研究生 郭晓薇

导师 张铁山

(北方工业大学经济管理学院 北京 100144)

摘 要 本文以食品冷链物流企业为研究对象,通过对食品冷链物流企业低碳运营现状的分析,发现了其在运营过程中存在的问题与不足之处。并引入具体案例,分别运用价值工程的基本公式和扩展公式对其进行价值分析,然后针对两者价值分析结果不一致的情况,运用寿命周期成本与碳排放量构建的评价模型,建立了综合比较模式。在综合比较模式下进行价值分析,寻求其有待改进和优化的突破口,从而为食品冷链物流企业有效地开展低碳运营及提高其经济效益与社会效益提供了参考与建议。

关键词 价值工程;食品安全;冷链物流;低碳运营

作者简介 郭晓薇,女,1989年出生,硕士研究生,主要研究方向为技术创新与价值工程。

E-mail: bw.gxw.19891108@163.com

1 绪论

1.1 研究背景与研究意义

随着消费者对食品安全重视程度的不断增强,冷链食品的市场迅速扩张,每年冷链食品的产量都会再创新高。冷冻冷藏食品市场的发展速度快,2013年高达9200万吨冷链市场需求,更加验证了我国冷链市场的巨大潜力。我国政府更是颁布了一系列政策大力支持冷链物流的发展。十八大明确提出要大力推进生态文明建设,坚持节约资源和保护环境的基本国策,着力推进低碳发展。2015年李克强总理在政府工作报告上明确指出,今年CO₂排放强度要降低3.1%以上。2013年,我国物流产业占能源消耗超过7%以上,给环境带来了严重污染。食品工业在各个环节都要消耗掉大量能源。冷链成本居高不下,严重阻碍了我国冷链物流产业的发展,通过引入价值工程的方法和理论可以有助于同时降低成本与碳排放量。

1.2 食品冷链物流企业低碳运营现状

由于我国食品冷链物流现阶段仍处于初级阶段,因此许多食品冷链物流企业普遍存在运营成本居高不下、设施陈旧、技术落后等诸多问题。通过调查,在我国,尚存在一些食品冷链物流企业出于成本的考虑,闲置先进的冷链设施以及冷链技术,并且缺乏低碳运营管理的意识与理念的现象尤为严重。主要表现为以下六个方面:一是低碳物流的发展政策相对匮乏;二是过于泛化的物流行业碳排放统计指标;三是完整的食品冷链物流体系尚未形成;四是冷链物流技术与设备闲置;五是食品安全意识淡薄;六是供应链脱碳意识缺乏。

2 相关理论综述

2.1 食品冷链物流概述

在生产、储藏、运输、销售直到最终到达消费者手中的各个环节中,使食品始终处于规定的低温环境下,以保证食品质量,减少食品损耗的一项系统工程,即为食品冷链物流。其特点如下:一是食品安全是首要目标;二是时间敏感性;三是高成本;四是信息多而复杂。食品冷链物流由冷冻加工、冷冻储存、冷藏运输与配送、冷藏销售及冷链逆向物流五部分构成。

2.2 价值工程概述

价值工程起源于 20 世纪 40 年代,由美国通用电器公司的麦尔斯(L. D. Miles)创立,并很快得到了推广。在美国价值工程已经形成了较完善的应用机制以及法律环境,政府法律规定超过 200 万美元以上的项目,尤其政府投资的公共项目或基础设施必须实施价值工程,以控制项目成本。日本在 20 世纪 50 年代引入了价值工程,之后,又传到了欧洲的一些国家。价值工程在世界各国的发展大致经历了四个阶段,即降低材料费用、改进现有产品、新产品的价值工程和系统的价值工程阶段。世界各国的价值工程组织,在 1959 年美国创建价值工程师协会以来已达 40 个。现在价值工程已经涵盖到了工业、建筑业、运输业、卫生保健、政府和环境等很多领域。超过 2/3 的世界 500 强公司持续开展价值工程活动。

2.3 低碳经济理论概述

21 世纪初,《我们能源的未来:创建低碳经济》在英国颁布后,诞生了低碳经济这一概念,而 1992 年的《联合国气候变化框架公约》则为低碳经济较早的的相关研究。20 世纪 90 年代,《联合国气候变化框架公约》与《京都议定书》相继出现,意味着联合国已对温室气体有了高度的重视,通过 150 多个国家的共同参与,决心一起努力减少温室气体,还原世界一个良好的生活系统,并明确制定了温室气体排放的目标、各国承担的责任与实现机制。21 世纪后,全球开始制定出更为细致的策略与方案,以面对日益恶化的空气环境。为了应对气候的变化,联合国制订了“巴厘岛路线图”,并要求发达国家在 2020 年前将温室气体降至 25%—40%。次年,各国领导人在 G8 峰会上也纷纷表示,在 2050 年实现全球温室气体排放减少 50% 的长期目标。2009 年我国在《中国发展低碳经济途径研究》一文中将低碳经济定义为“一个新的经济、技术和社会体系,与传统经济体系相比在生产和消费中能够节省能源,减少温室气体排放,同时还能保持经济和社会发展趋势的发展模式”。

3 基于 VE 基本公式的食品冷链物流企业的价值分析

3.1 食品冷链物流企业的运营

食品冷链物流企业运营的目标主要包括:一是服务好;二是快速及时;三是成本低;四是质量高。食品冷链物流企业运营主要包括以下基本内容:预测食品冷链物流的需求、设施选址、管理冷链物流设施及设备、管理冷链物流运输、管理冷链物流相关组织、订单管理及客户服务、管理冷链物流库存、管理冷链物流信息系统、控制冷链物流成本以及控制冷链物流质量十个方面。

3.2 运用 VE 基本公式进行食品冷链物流企业的价值分析

通过食品冷链物流的定义以及构成,可以明确冷链是三位一体的,三位即设备、人员和流程,一体即通过对三者的管理而形成的冷链系统。因此食品冷链物流企业运营过程的基本功能是保障食品品

质,按照形成冷链流程、提供冷链设备和技术以及保证人员三个方面可划分出保证温度、提供包装、利于节能、提高效率、便于操作、完善服务、提高熟练度等十二个辅助功能,并据此建立了功能系统图。

本文以 C 企业为案例,分别邀请 VE 小组成员 2 名、食品冷链物流企业的管理人员和操作人员各 3 名对食品冷链物流企业运营过程的功能进行功能评价,通过“0—4 两两对比法”,最终得出功能系数。根据食品冷链物流企业运营过程中在各功能环节上耗用的物资、能源等数量的比例,对食品冷链物流企业运营的成本进行分摊,从而确定出成本系数。通过计算可以得到食品冷链物流企业运营过程的价值系数,如表 3—1 所示。

表 3—1 食品冷链物流企业运营过程价值系数表

功能名称	功能系数(%)	成本系数(%)	价值系数 $V_i = F_i / C_i$
保证温度	11.36	17.04	0.67
提供包装	10.61	9.49	1.12
完成加工	9.28	8.76	1.06
装卸搬运	8.95	12.30	0.73
处理信息	7.53	7.94	0.95
利于节能	9.00	10.24	0.88
提高效率	8.71	4.57	1.91
便于操作	6.96	6.24	1.12
完善服务	7.00	4.13	1.69
提高熟练程度	7.53	5.77	1.31
避免不当操作	8.24	8.34	0.99
加强激励	4.83	5.20	0.93

4 基于 VE 扩展公式的食品冷链物流企业的价值分析

4.1 食品冷链物流企业碳排放特征

经过企业能耗调查,总结归纳出食品冷链物流企业的主要碳排放特征如下:一是冷库是食品冷链物流企业中的主要用电设备,因此也是节能的核心设备,冷库节能最为重要;二是企业根据用电波峰波谷价格差异,尽量使用低谷电,采取“峰段用电尽量不使用,尖峰用电坚决不使用”的原则;三是低温库能耗比高温库耗能要小得多;四是冷链运输过程中碳排放的来源主要是行驶耗油。

4.2 食品冷链物流企业的碳排放量测算

通过对比国际碳排放核算标准以及碳排放量的五种测算方法,以及食品冷链物流企业运营过程中碳排放的特点,确定其采用生命周期评价法测算碳排放量较为适宜。根据 PAS2050 的规定及生命周期评价法,计算食品冷链物流企业碳排放量的基本步骤如下:第一步,建立食品冷链物流企业的运营流程图;第二步,对食品冷链物流生命周期的边界进行界定以及确定优先序;第三步,收集并计算数据;第四步,各流程碳排放量的计算公式为:某流程碳排放量=碳排放系数×动态数据。

4.3 食品冷链物流企业碳排放的影响因素研究

冷库所需的冷量,主要有以下几个方面组成:建筑围护结构的漏热量、食品进出库的总温差、操作

总热量、通风换气热和照明热量等。在食品储藏期间,建筑围护结构的漏热量占相当的比例;食品进出冷库时的总温差与冷库的实际使用要求有关;操作总热量与冷库的设计操作方法有关;通风换气热主要与开门的次数和对流换气量有关;照明热量主要与库房面积和光照度及灯具的形式等有关。在冷库所需总冷量确定的前提下,冷库的能耗仅与制冷设备的效率有关。降低冷库的能耗,可以从减少冷库所需的总冷量和提高制冷系统单位输入电耗所获得的制冷量的角度来考虑。

4.4 运用 VE 扩展公式进行食品冷链物流企业的价值分析

根据 C 食品冷链物流企业运营过程中实际的碳排放量、按功能要素划分的食品冷链物流企业的碳排放情况表的分配情况,以及经过与 VE 小组专家、食品冷链物流企业主要设备与能源管理人员共同协商确认后,得到了食品冷链物流企业运营过程碳排放量功能分摊比例,确定出碳排放系数。从而计算得到扩展公式下的食品冷链物流企业运营过程价值系数,如表 4-1 所示。

表 4-1 扩展公式下的食品冷链物流企业运营过程价值系数

功能名称	功能系数(%)	碳排放系数(%)	碳排放价值系数 $V_i = F_i / C_i$
保证温度	11.36	13.85	0.82
提供包装	10.61	13.28	0.80
完成加工	9.28	7.54	1.23
装卸搬运	8.95	11.85	0.76
处理信息	7.53	7.64	0.99
利于节能	9.00	7.82	1.15
提高效率	8.71	5.08	1.71
便于操作	6.96	5.48	1.27
完善服务	7.00	6.88	1.02
提高熟练程度	7.53	6.15	1.23
避免不当操作	8.24	8.53	0.97
加强激励	4.83	5.92	0.82

5 基于综合比较模式提升食品冷链物流企业价值的途径

5.1 基本公式与扩展公式价值分析的综合比较

两种价值分析得出的结果相比,保证温度、装卸搬运、处理信息、避免不当操作与加强激励功能的价值系数均小于 1,说明它们为重点的改进对象;而完成加工、提高效率、便于操作、完善服务与提高熟练程度功能的价值系数均大于 1,碳排放量与成本无需做较大改变,只需验证是否实现功能与有无过剩功能,若有,则剔除过剩功能以降低成本与碳排放量。提供包装与利于节能两个功能在价值工程基本公式、价值工程扩展公式的价值分析后得出的结论截然相反,为使结果准确且一致,需建立综合比较模式,因此需搭建两者相关联的桥梁。

5.2 基本公式与扩展公式综合比较模式的构建

(1)两者的价值分析结果一致。意味着在采用某项措施改进某功能时,寿命周期成本与碳排放量的改变方向一致。若两者价值系数均大于 1,意味着此时的寿命周期成本与碳排放量比较适宜,无需

做较大的变动。若两者价值系数均小于 1,意味着此功能是重点改进对象,需分别从寿命周期成本、碳排放量两个角度考虑改进措施。

(2)两者的价值分析结果不一致。两种公式得出功能的价值系数出现矛盾,对同一功能采取相同的改进措施,可能造成 VE 基本公式与扩展公式的价值系数出现相反的情况。因此,引入了碳税成本与碳交易收益的概念,将碳排放量转化为可量化的成本进行核算,使其与寿命周期成本变为同一概念。此时,若某功能的价值系数仍小于 1,便可以列为重点的改进对象,若价值系数大于 1 或者小于 1,基本上可以不做较大的改动或不改动。

5.3 运用综合比较模式的食品冷链物流企业的价值分析

利用综合比较模式进行价值分析的方法与步骤如图 5-1 所示:

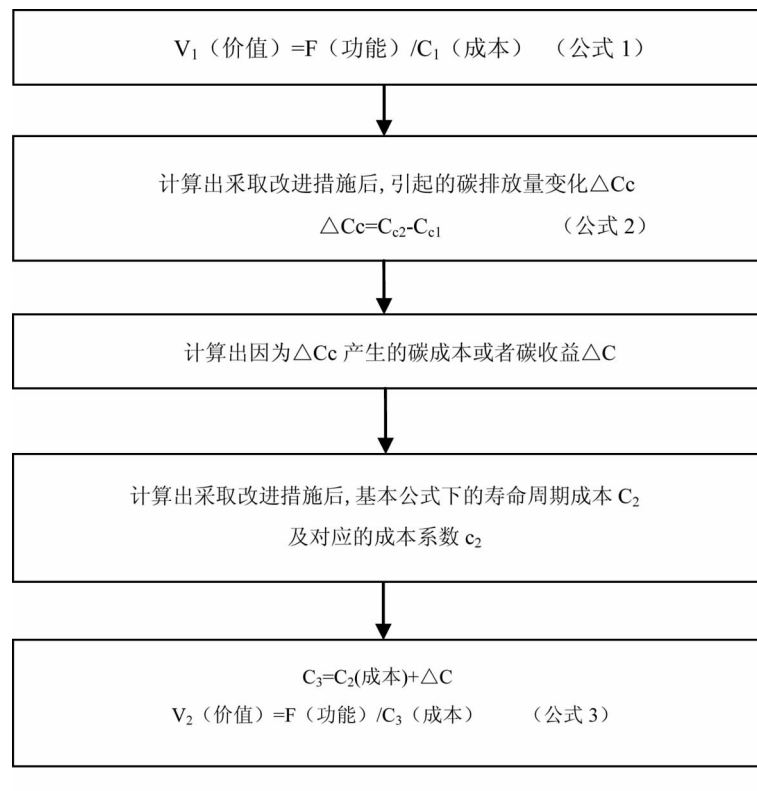


图 5-1 综合比较模式下价值分析的方法与步骤

(1)提供包装功能在综合比较模式下的价值分析。在价值工程扩展公式下,采用可降解的薄塑料进行食品包装,可相应降低 35%的碳排量,使用阶段碳排放量综合减少 7.07 吨,以 2013 年北京市全年碳排放权交易均价为 50.2 元/吨计算可知,每年更换包装材料在寿命周期内碳收益为 $7.07 \times 50.2 = 354.914$ 元。但使用该种材料,每年购买包装材料的成本将增加 2.7 万元,故包装成本为 $(154.66 + 2.7)$ 万元,即 157.36 万元。

$V_2 = 0.1061 / [(1573600 - 354.914) / 16290000] = 1.09 < 1.12$ 。此方法行不通。若食品冷链企业不愿意采取该项措施进行低碳减排也是可以理解的,政府为了号召企业节能减排,应该为愿意实行此措施的企业发放一定的现金补贴或优惠政策,以推进企业低碳运营,从而实现双赢的局面。

(2)利于节能功能在综合比较模式下的价值分析。在价值工程扩展公式下,可通过使用计算机自

动控制冷库温度进行节能,可相应降低 25% 的碳排放量,使用阶段碳排放量综合减少 6.15 吨,按照 2013 年北京市全年碳排放权交易均价为 50.2 元/吨计算可知,每年用计算机自动控制冷库温度在寿命周期内碳收益为 $6.15 \times 50.2 = 308.73$ 元。但使用计算机,每年冷库成本将增加 200 元,故成本为 1667620 元。

$V_2 = 0.09 / [(1667620 - 308.73) / 16290000] = 0.88 = V_1$ 。故此项措施可以实施。

5.4 综合比较模式下提升食品冷链物流企业价值的优化方案与建议

(1) 提升保证温度功能价值的优化方案与建议。使用时间—温度指示器(Time-Temperature Indicators, TTIs)以实现在各个节点能够控制食品冷链。可以通过指示器的颜色渐变过程直接推断食品的新鲜程度和剩余存储期。在欧盟项目 SMAS(安全监测和保障体系)范围内测试了 TTIs 的使用,通过此项目,质量不合格的食品可以从 12.5% 降低到 4.32%。租赁 C 企业冷库的 K 企业,与 C 企业的性质不同,是一家外资企业,冷链设备设施与技术都较为先进,尤其对监控温度的设备设施与技术更为重视。整个冷链过程都采用了时间—温度指示器与温度控制系统,正好与 C 企业形成较为鲜明的对比。

(2) 提升装卸搬运功能价值的优化方案与建议。在装卸搬运的过程中,应注意到食品易损伤易腐蚀等特点,在放有集装箱的地板上需要铺上一定的覆盖层,及时更换已损坏的缓冲冲气囊,以及在托盘间填充相应材料以保障集装箱内部的温度和湿度。C 企业在实行以上优化方案一段时间后,大大避免了在装卸搬运过程中造成的损失与浪费,同时节约出了可观的成本以及明显缓解了环境污染情况,经过统计与计算,成本系数与碳排放系数减少幅度分别为 4.17%、1.01%。

(3) 提升处理信息功能价值的优化方案与建议。立足当前冷链物流领域的基础,依靠云计算、物联网、智慧物流等先进技术手段,升级改造食品冷链物流企业的信息化系统,从而加强“全程透明”冷链的控制管理。以互联网的思想推动餐饮冷链、生鲜宅配等市场的开拓与创新。激励行业内和区域间的冷链物流平台信息的共享,增进冷链物流信息互联互通。

(4) 提升避免不当操作功能价值的优化方案与建议。首先应进行初步调查以确定食品的损害来源,然后采取以下改进措施:使用较坚韧的包装材料,避免底部的货箱被上层货箱压坏;充分保证空气流动,避免食品的本身温度高于制冷系统的温度;定期检查制冷设备,防止发生故障导致温度不稳定变化;集装箱中合理的混合贮存,可以将贮存条件类似的食物放入同一集装箱内,但不要将贮存条件差异较大的食物一同存放。C 企业进行了为时一个月的比较实验,一个月后发现,经过培训的专业冷链操作人员都采取了较为准确的冷链操作行为,果蔬剩余质量较多且保存较完好,而且完成效率较高;而采取不当操作的普通冷链操作人员,果蔬损失较为严重,而且效率低下。

(5) 提升加强激励功能价值的优化方案与建议。首先,检测出各部门能源消耗以及成本消耗的平均值,从而对进行开源节流的员工奖励有评定依据;其次,鼓励企业上下各阶层员工,共同参与企业开源节流活动中去,积极思考降低成本与减少碳排放量的方法与对策;最后,对于为实现企业开源节流目标做出贡献的员工做出相应的奖励,以激励全体员工共同努力。C 企业听取了笔者的建议后,针对员工建立了较为详细的奖惩条例,并严格地执行了一个月后发现,很多员工意识到企业的经济效益与社会效益都关乎自己的利益,便开始为企业开源节流提出了很多对策与建议,并改变了以往对于企业浪费现象置之不理的态度,主动去帮助企业共同降低企业的运营成本与能源消耗量。

6 结论

通过以上研究与分析,可以得到以下结论:(1)食品冷链物流运营过程的基本功能为保障食品品

质;(2)适合于食品冷链物流企业运营过程中各阶段碳排放测算的方法为生命周期评价法;(3)可通过引入碳税成本与碳交易收益的概念,建立了寿命周期成本与碳排放量的评价模型,从而得出经济效益和社会效益兼顾的对策与建议,以实现食品冷链物流企业低碳运营的目标;(4)验证了价值工程、食品冷链物流与低碳相结合的可行性以及具有良好的收益效果。

参 考 文 献

- [1] 刘倩晨. 考虑碳排放的冷链物流研究[D]. 北京:清华大学,2010
- [2] 田丽娜. 考虑碳排放的水果单级冷链效益优化分析[D]. 成都:西南交通大学,2012
- [3] 白世贞,曲志华. 冷链物流[M]. 北京:中国物资出版社,2012
- [4] 毋庆刚. 我国冷链物流发展现状和对策研究[J]. 中国流通经济,2011,(2):24—28
- [5] 兰洪杰,康彪. 北京市食品冷链物流问题探析[J]. 物流科技,2010,(6):7—10
- [6] Richard S. J. Tol. Carbon dioxide emission scenarios for the USA [J]. Energy Policy, 2007, (35): 5310 - 5326

The Application of Value Engineering in Food Cold-chain Logistics Enterprises for Low-carbon Operation

Guo Xiaowei

(School of Economics and Management, North China Univ. of Tech. , Beijing, 100144)

Abstract The research object of this paper was the food cold-chain logistics enterprises, through the analysis of the current situation of low-carbon operation in food cold-chain logistics enterprise, found the existing problems and deficiencies in the process of operation. And introduced the concrete case to analyze the value of basic value engineering formula and extension value engineering formula, then using the evaluation model of life-cycle cost and carbon emissions if there are both not consistent, to build up comprehensive comparison model. Analyzed the value of comprehensive comparison model, to seek its improvement and optimization, and to provide a reference suggestion for food cold-chain logistics enterprises to carry out low-carbon operation effectively and improve the economic benefits as well as social benefits.

Key words value engineering; food safety; cold-chain logistics; low-carbon operation

强制医疗责任保险立法问题研究

研究生 钟 蕾

导师 王 瑞

(北方工业大学文法学院 北京 100144)

摘 要 近年来层出不穷的医疗纠纷甚至医闹事件备受关注,医疗机构和患者的利益都亟需得到保护。医疗责任保险既能转嫁医疗机构的医疗风险又能及时保护患者利益,本应该得到广泛应用并发挥其应有作用。然而,目前医疗责任保险在我国却发展缓慢,医疗责任保险缺乏法律层面的强制,而行政主导方式又存在诸多弊端,不利于医疗责任保险的健全发展。因此,本文提出在立法上对医疗责任保险进行强制,在立法上建立全面系统的医疗责任保险制度,不仅可以分散化解医疗风险,更关键的是可以保护患者利益。这也是为满足保险大数法则的要求。此外,这也是我国医疗卫生体制改革的必然要求。在立法的可行性方面,不仅可以借鉴国外经验,而且我国也有医疗责任保险试点经验以及机动车交通事故责任强制保险制度的经验,再加上对我国医疗责任保险市场重视和开拓,因此强制医疗责任保险在立法上是可行的。但不可否认也存在如数据、人才等方面的立法障碍。在全面分析上述内容的基础上,笔者将提出立法建立强制医疗责任保险的相关建议。

关 键 词 医疗纠纷;医疗责任保险;强制医疗责任保险

作者简介 钟蕾,女,1990年出生,硕士研究生,主要研究方向为民商法学。

E-mail:cszhonglei@126.com

1 绪论

保险的风险转移作用是众所周知的,强制医疗责任保险,以立法的形式强制规定参加医疗责任保险是医疗机构和医务人员的法定义务,以此来实现分散医疗风险和及时保护受害患者的目的。本文的分为四个部分:我国医疗责任保险立法现状、强制医疗责任保险立法的必要性、强制医疗责任保险立法可行性及障碍以及强制医疗责任保险的立法建议。医疗责任保险关乎医疗机构的利益、医务人员的利益,更关系到患者利益,其重要性不言而喻。正因如此,在立法时更要慎重考虑,不能操之过急,更不能为立法而立法,应该进行反复再三的论证,以使所立之法在符合我国国情的基础上真正发挥作用。

2 正文

2.1 我国医疗责任保险立法现状

《中华人民共和国保险法》(以下简称《保险法》)第65条第4款规定“责任保险是指以被保险人对第三者依法应负的赔偿责任为保险标的的保险。”强制医疗责任保险,属于职业责任保险中的一种,是指在法律层面规定参加医疗责任保险是医疗机构或者医务人员的一种强制的义务,这种保险是国家立法强制实施的,不仅是医疗机构和医务人员的一种义务,甚至可以作为医疗机构和医务人员取得相应资质或资格的必要条件。

2.1.2 立法空白

目前,我国关于医疗责任保险相关的规定主要是规范性文件。根据《保险法》第11条第2款“除法律、行政法规规定必须保险的以外,保险合同自愿订立。”这些规范性文件不属于法律意义上的强制保险,属于由政府主导并推动的政府统保模式,简称为“行政主导模式”。因此,强制医疗责任保险在我国的立法处于空白状态。

2.1.3 行政主导模式的弊端

行政主导模式最大的弊端就是强制性不足。行政主导下的医疗责任保险,更多的是行政力量的作用,而不是医疗机构的自发行为,在这种状况下,出现的结果只能是投保数量的不断下降以及医疗责任保险的低迷发展,医疗机构对参加这一保险没有兴趣甚至有些抵触,保险公司也对这一领域抱着消极的态度。

2.1.4 立法强制方式的提出

本文主张对医疗责任保险进行立法,在法律上明确医疗责任保险是一种强制保险,对保险当事人及其权利义务、运作方式、保险范围等做出明确的规定,并不断健全医疗责任强制保险的相关内容,利用法律的规范性、权威性及强制性将对医疗机构的外在强制转化成为其自觉的内在动力,从而拉动市场对医疗责任保险真正的、持续的需求,避免“行政主导模式”运行中出现的弊病。

2.2 强制医疗责任保险立法的必要性

2.2.1 分散医疗风险的需要

近年来,医疗纠纷不断增加,各种医闹事件层出不穷,医患矛盾一度成为社会焦点并逐渐演变为一个不容忽视的社会问题。对于医疗损害责任纠纷案件,长期以来司法实践中的三个双轨制^①的做法无疑加大了医疗诉讼中医疗机构的压力,也加大了医疗风险。另外医学活动中不可避免因患者个体差异等因素带来的治疗风险也是医疗风险的一部分。强制医疗责任保险的确立,可以从实质上帮助医疗机构分散化解上述医疗风险。

2.2.2 保护患者利益的需要

侵权损害赔偿的救济存在一些弊端:第一,救济程序复杂、及时性较差。第二,救济范围较小。第三,举证责任分配和医疗损害鉴定的不完善,第四,医疗诉讼的结果具有不确定性。在医疗损害案件中,患者的利益亟需保护。另外,目前普遍存在的防御性医疗行为,毫无疑问损害的最终是患者的利益。

2.2.3 保险大数法则的需要

大数法则作为保险经营的根本准则,跟切比雪夫大数定理、伯努利大数定律以及辛钦大数定律密不可分。上述三个定律是大数法则的来源和基础,其原理概括如下:随机现象在相同条件下不断重复,当其重复率达到一定量时,出现的结果就会有规律可循,而且这样的规律比较稳定。在医疗机构投保数量不足的情况下,保险公司不愿承保。因此,在立法上规定医疗机构和医务人员必须投保,以增加投保的数量,满足大数法则的在医疗责任保险中运行的需求。

^① 第一个双轨制是医疗事故与医疗过错的双轨制,第二个双轨制,是指在赔偿时,如果认定是医疗事故责任的,法院以《医疗事故处理条例》为依据确定相关赔偿事项,而确定为医疗过错责任的,法院则适用人身损害赔偿的标准来确定赔偿标准。第三个双轨制体现在医疗鉴定中,即在医疗鉴定中,既可以由医学会做出鉴定,也可以由司法鉴定机构做出鉴定。

2.2.4 医疗卫生体制改革的需要

身体的健康与否直接关系到每一个人生活的幸福不幸福,如何让患者放心地就医、如何让医疗机构和医务人员安心投入到医疗活动当中,如何缓和医患关系,妥善处理医疗纠纷甚至医闹事件,目前在我国卫生医疗体制改革中至关重要。

2.3 强制医疗责任保险立法的可行性及障碍

2.3.1 强制医疗责任保险立法的可行性

第一,国外经验借鉴。他山之石,可以攻玉。西方发达国家如美国、英国、德国以及法国等基本上都是实行强制医疗责任保险,大多都在立法中将投保医疗责任保险作为医生执业条件之一,医生出于保护自身利益也会积极投保这一险种。在今后我国医疗责任保险发展中,需要认真学习和总结国外医疗责任保险的做法和经验,取长补短,不断地推动我国医疗责任保险稳步向前发展。

第二,我国试点地区经验总结。云南、上海、北京以及深圳等地进行医疗责任保险试点探索。通过对四个地区开展医疗责任保险试点情况的简介,主要特点如下:第一,根据医院性质,强制非营利性医疗机构投保,其他医疗机构自愿参加。第二,承保范围逐渐扩大。第三,成立专门的调解或鉴定机构。

第三,机动车交通事故责任强制保险经验借鉴。《机动车交通事故责任强制保险条例》,是2006年3月21日颁布的,在同年7月1日起生效。“交强险”提高了交通事故投保率、保障受害人迅速获得赔偿、纠纷得到快速有效地处理、维护了正常的交通秩序。医疗行业和道路交通安全行业类似,都属于高风险行业,因此,医疗责任保险的立法应该认真研究上述条例,借鉴其中费率制定、险种设计、交通事故社会救助基金等诸多成熟完整的运作模式。

第四,医疗责任保险市场广阔且逐渐得到重视。从我国来看,我国医疗责任保险市场发展潜力不可小觑。截至2014年11月底,我国医疗卫生机构数达98.5万个。目前保险公司以各医疗机构的床位数为依据计算医疗机构所需要缴纳的保费,床位数在301—400的,年度保险费4.5万元。^①以北京市为例,^②平均每个公立医院的床位数约是329张^③,那么每个公立医院的年度保险费就是4.5万元,总的保费就是1228.5万元。如果只有以100家公立医院投保,需要达到在所有公立医院投保下的总保费,每个公立医院需要缴纳的保费就是12.285万元,每个公立医院需要多交7.785万元。这样的简单估算还只是以公立医院和民营医院为基础,如果再加上其他医疗机构,医疗责任保险市场的潜力就更大了。再加上近年来,国家陆续出台医疗责任保险的相关规定,该保险越来越受到重视。如2014年7月,国家卫生计生委会同司法部等部门联合发布了《关于加强医疗责任保险工作的意见》。

2.3.2 强制医疗责任保险的立法障碍

第一,我国医疗责任保险制度不完善。我国《保险法》关于责任保险的规定比较笼统,实践中运行时困难重重,责任保险本身也是我国保险法领域亟需改进完善的地方。如在患者直接请求权方面没有规定以及保险人对医疗事故的处理过程也没有抗辩义务。

第二,我国医疗责任保险缺乏数据、人才。目前保险公司缺乏医疗机构关于风险评估和费率厘定

^① 上述数据来源《上海市医疗事故责任保险实施方案(试行)》,参见吕群荣. 医疗责任保险制度法理基础与制度构建[M]. 北京:中国政法大学出版社,2014.

^② 北京市目前有各类医疗机构10141个,其中医院647个(公立医院273个、民营医院374个),基层医疗卫生机构9261个,专业公共卫生机构121个以及其他机构(疗养院、医学科学研究机构等)121个。2013年,北京市医疗机构床位数总计是116801张,其中公立医院89785张,民营医院16761张。<http://www.phic.org.cn/tonjixinxi/>,北京市卫生计生委官网,2013年北京市卫生工作统计简编,2015年3月25日访问。

^③ 公立医院的总床位数除以公立医院的个数。

的基础数据,如与各医疗机构内部管理及经营水平以及各科室和医务人员过去发生医疗损害的历史数据等数据。目前保险公司缺乏这种既有医学知识和技术,又懂保险和法律的复合型人才。

第三,对医疗责任保险的认识存在误区。很多医疗机构对医疗责任保险的认识存在误区,如一些医疗机构认为购买医疗责任保险,会使患者认为是医疗机构“心虚”的表现,是对自己医疗技术不自信的表现。此外,很多人认为卖保险就是行骗,保险意识不强。

2.4 强制医疗责任保险的立法建议

2.4.1 立法目的

在立法目的上应强调以下几个方面的内容:首先,保障医疗机构以及医务人员的合法利益。第二,保护患者的合法利益。第三,鼓励医务人员提高自身医学知识、治疗水平以及服务质量,促进医学发展和创新。

2.4.2 立法内容

第一,投保人、被保险人以及保险人。投保人不仅包括医疗机构,还包括医务人员。根据《侵权责任法》第54条的规定,医疗损害责任的责任主体是医疗机构,因而医疗机构是该保险的被保险人。在今后立法时,应当对从事医疗责任保险的保险公司设置一些硬性条件,同时,保监会也要加强对保险公司从事医疗责任保险的指导和监督。

第二,合同主体的权利义务。在立法时应当确立下面几个方面的内容:首先,保险人的索赔参与权。其次,受害人直接请求权。再次,医疗告知义务。

第三,保险责任范围与免责范围。在今后对医疗责任保险进行立法时,在责任范围的拟定时,应当严格以《侵权责任法》为根本的法律依据来界定赔偿责任。另外,下面两个方面也应纳入承保范围:(1)见习生、实习的医学院校学生因过失造成的损害。(2)精神损害赔偿。立法时,对于免责范围应该做出严格明确的要求。

第四,保险费率的厘定。当前保险公司在厘定费率时,应该考虑以下几方面,第一,认真参考借鉴“交强险”。第二,具体问题具体分析,各个医疗机构的保险费率是根据各医疗机构的具体情况厘定的。第三,在厘定费率时,必须考虑各个医疗机构以及医务人员参加医疗责任保险前,该医疗机构的医疗损害记录,以及医务人员的执业经历。

第五,保险期限。立法时应明确采取“期内索赔式”,并且合理制定“追溯期”制度与“延长报告期”制度。

第六,保险金额。最低限额方面,采取“单笔索赔自负额”。最高限额设定时,不仅结合各地区医疗水平等基本情况,而且要充分征求保险公司、医疗机构以及患者的意见。

第七,医疗纠纷调解处理机构。医疗纠纷调解处理机构要顺利开展纠纷调解处理工作,在立法时需要明确“三性”即独立性、中立性和专业性。

第八,医疗损害鉴定机构。只有真正准确、权威、专业、公正的鉴定才能使医疗机构与患者之间的争议平息。在立法时还应当详细规定医疗损害鉴定的相关内容,如医疗损害鉴定的程序、鉴定结论的出具以及错误鉴定甚至徇私枉法鉴定的惩处措施等等。

第九,医疗事故汇报制度。医疗事故汇报制度是通过建立一个医疗损害案件的信息交流平台,形成一个稳定的全面的医疗损害案件信息系统。

2.5 结论

要改变我国医疗责任保险的现状,目前最急需的在法律层面规定医疗责任保险是一种强制保险,并构建统一的强制医疗责任保险法律制度。只有通过法律确立的统一的强制医疗责任保险制度,才能

为医疗责任保险良好运行保驾护航。

参 考 文 献

专著:

- [1] 姜凤武著. 医疗损害责任制度比较研究[M]. 北京: 法律出版社, 2013. 180—191
- [2] 杨静毅著. 医疗侵权的经济分析[M]. 北京: 法律出版社, 2013. 208—220
- [3] 蒲川, 冯磊著. 医疗纠纷非诉讼处理机制研究[M]. 重庆: 西南师范大学出版社, 2012. 112—162
- [4] 曾言, 李祖全著. 医疗责任强制保险制度研究[M]. 长沙: 湖南师范大学出版社, 2009. 114—173
- [5] 李青武著. 机动车责任强制保险制度研究[M]. 北京: 法律出版社, 2009. 12—57
- [6] 吕群蓉. 医疗责任保险制度法理基础与制度构建[M]. 北京: 中国政法大学出版社, 2014

期刊论文:

- [7] 张泽洪. 论医疗责任保险的强制[J]. 中国保险, 2010, (10): 22—24
- [8] 李海澈. 强制医疗责任保险的政府属地化干预及法律控制[J]. 医学与社会, 2013, 26(7): 77—79
- [9] 翟宏丽. 我国医疗责任强制保险的法律定位[J]. 医学与法学, 2012, 04(4): 15—16
- [10] 沈庆劫. 信息不对称视角: 我国医疗责任保险实施强制立法之管见[J]. 现代财经, 2011, 31(2): 46—50
- [11] 吕罗荣, 刘超婷, 罗德军等. 以强制医疗保险为主的医疗纠纷保障模式探讨[J]. 解放军医院管理杂志, 2012, 19(2): 163—165
- [12] 董文勇. 我国医疗责任保险法律制度构建的问题与方案[J]. 河北法学, 2014, 32(6): 142—148

博硕士论文:

- [13] 沈思言. 医师专家责任保险研究. 山东大学硕士学位论文, 2005(3)
- [14] 任婷. 中外医疗责任保险法律制度的比较研究. 中国海洋大学硕士学位论文, 2007(6)
- [15] 胡慧芬. 强制性责任保险制度研究. 华东政法学院硕士学位论文, 2006(11)
- [16] 李莎. 我国医疗责任保险模式研究. 对外经济贸易大学硕士学位论文, 2012(5)
- [17] 谭亭. 重庆市实施强制医疗责任保险的立法可行性研究. 重庆医科大学硕士学位论文, 2009(5)
- [18] 张磊. 中国强制责任保险制度研究. 厦门大学博士学位论文, 2007(5)

外文资料:

- [19] Patrick T. O'Rourke, Kari M. Hershey [EB/OL]. http://www.the-hospitalist.org/details/article/233905/The_Power_of_sorry.html
- [20] Marc Stauch, The Law of Medical Negligence in England and Germany : A Comparative Analysis[M]. Oxford, 2008: 146—147

Research on the legislation of the compulsory medical liability insurance

Zhong Lei

(School of Humanities and Law, North China Univ. of Tech., Beijing, 100144)

Abstract The frequent occurrences of medical disputes have attracted widely attention. The doctor-patient relationship is becoming more and more nervous. Both the interests of the medical in-

stitution and the patient need to be protected. The medical liability insurance can achieve this purpose. So in this paper, the author suggests establishing the compulsory medical liability insurance. It can not only dissolve the medical risks, but also can protect the interests of patients. It also is the inevitable requirement of law of large numbers and the reform of medical and health system. In the process of establishment, both can draw lessons from foreign experience and the pilot experience of medical liability insurance and the experience of the compulsory traffic accident liability insurance for motor vehicles. But there is no denying that there are still some obstacles of legislation. On the basis of above analysis, the author will put forward some legal suggestions.

Key words medical disputes; medical liability insurance; compulsory medical liability insurance

网络失范言论的刑法規制

研究生 曾 磊

导师 吴焱光 王海桥

(北方工业大学文法学院 北京 100144)

摘 要 信息时代,网络失范言论导致的违法犯罪现象日益增多。尽管公民言论自由受到宪法保障,但这种保障也是有限度的。一旦滥用就会侵犯人权,甚至破坏社会秩序和国家安全。在我国当前宽严相济刑事政策指导下,应对网络失范言论的刑法規制予以区分对待。针对我国刑法中可以规制的相关罪名及其产生的司法疑难,本文在借鉴域外法的基础上,主张通过刑法修正案、出台司法解释以及严格使用刑事司法等方式对其加以完善。

关键词 网络失范言论;言论自由;刑法边界;規制分析;完善建议

作者简介 曾磊,男,1990年出生,硕士研究生,主要研究方向为刑法学。

E-mail: lyfzzll@163.com

1 绪论

网络技术的发达使得民众已变成了信息的制造者和传播者,公民的言论自由得以彰显。然而,一系列网络失范言论事件却导致了网络环境的污浊,严重破坏了正常的社会秩序。我国网络言论失范与否的界定标准未得到统一论,现有成文法难以对其进行全面适用,因此,在司法过程中,出现了适用疑难问题。本文以网络失范言论为研究对象,对刑法上的相关罪名进行重点归纳,结合案例分析其存在的适用困难,最终通过对国外相关法律法规的分析和借鉴,提出完善我国网络失范言论刑法規制的建议,希望以此达到解决问题、平衡利益的目的。

2 正文

2.1 网络失范言论之界定

网络失范言论是指,网民违背了一定的社会秩序和所应遵循的言论准则要求,以及因为不适当地接触和使用互联网而导致言论偏差的情况,体现了个人言论自由与公共网络秩序间的矛盾关系。

2.1.1 网络失范言论的特性描述

2.1.1.1 隐蔽性

传统言论通过语言和文字进行交流,需要言论主体正面接触与互动,过程相对确定。而网络失范言论的传播是通过互联网构建的虚拟世界,行为人匿名的身份得以规避责任,使执法者难以查到行为人的真实信息,给刑事司法认定工作制造了障碍。

2.1.1.2 即时性

传统方式上信息的反馈需要经过层级上报与审核,因而准确性和真实性有待考究。互联网的出现,其公开性和易复制性导致网络言论的传播速度极快,提高了传播效率,但也会造成相关违法犯罪行

为的不良影响迅速扩散、难以控制。

2.1.1.3 广泛性

网络失范言论的传播表现形式多样,不仅是文字,也可以是其他一切被外人所感知的外部载体,其载体呈现复杂化和无纸化。同时受到不同言论主体思想的碰撞,网络言论内容具有多样性。一旦构成犯罪,社会影响难以估量。

2.1.1.4 交互性

互联网的出现使信息的传播方式由放射状变为网状,网络用户不再是单纯的信息使用者或提供者,而是两者的统一结合。这就决定了信息的多样化,促进了信息的传播和交流,但同时网络失范言论也会形成舆论趋同,被大量传播或下载,导致社会危害性不可控,威胁信息网络安全。

2.1.2 网络失范言论的刑法边界

2.1.2.1 宪法保障下的言论自由

言论自由作为一项基本人权,为国际人权公约和区域性人权公约所确认和保护。我国宪法中也对言论自由予以明确保障,但宪法保护也并非绝对的,当言论自由给他人、集体或国家利益造成损害时,要受到适当限制,因此,网络失范言论超越了言论自由的范畴,不受宪法保护。

2.1.2.2 刑事政策视域中的网络失范言论

我国当前宽严相济的刑事政策要求,在坚持对严重刑事犯罪进行严厉打击的同时,适度放宽对轻微犯罪的惩罚,更多地强调非犯罪化、轻刑化和非监禁化。对网络失范言论而言,应在充分保障言论自由的基础上,遵守刑法的谦抑性。

2.2 网络失范言论的现行刑法分析

2.2.1 侵犯人格权的刑法规制

2.2.1.1 侵犯名誉权的刑法规制

2.2.1.1.1 诽谤罪

网络诽谤是诽谤行为在空间上的延伸,因而也应纳入诽谤罪的规制范围中。两高专门针对该类行为于2013年出台了相应的司法解释,为治理网络诽谤之不法行为提供了法律依据。“捏造事实”,是指行为人无中生有地编造、杜撰虚假事实,这种虚假事实不是客观存在的,而是行为人出于主观上的恶意而进行虚构的。“散布行为”通常是指行为人在公众场合将虚假信息向不特定的多数人进行传播的行为。此外还通过列举制定了“情节严重”的标准。

2.2.1.1.2 侮辱罪

侮辱行为是以暴力或者其他方法,笔者认为互联网即为其他方式之一。网络侮辱更多时候体现的是一种精神暴力或心理暴力。网络侮辱能够使不特定的人或是特定多数人所知悉,因此符合侮辱罪的公然性。在对“侮辱”进行认定时,笔者主张同时考虑被害人的主观态度以及该侮辱行为的社会影响。

2.2.1.2 侵犯隐私权的刑法规制

我国对于隐私权的保护处于名誉权保护的范围之内属于言论自由的间接保护。《刑法修正案(七)》中增设了非法提供公民个人信息罪,为维护网络信息安全提供了法律依据。但笔者认为犯罪主体不应限于规定中的六类,而应将其扩大为一般主体。“个人信息”指能够识别公民个人身份或者涉及公民个人隐私的信息、数据资料。“非法提供”包括非法散布与传播。对于“情节严重”的认定,笔者认为应从社会危害性、犯罪实行行为的细节以及参照类似罪名等综合考虑。

2.2.2 破坏公共秩序的刑法规制

2.2.2.1 传播淫秽物品罪

“淫秽物品”的实质是诲淫性,不论其载体形式如何,只要符合诲淫性的实质标准,即可认定为淫秽物品,因此互联网也可以成为淫秽物品的传播载体,网络传播方式更为虚拟化和多样化。笔者认为上传行为不是结果犯而是抽象危险犯,一旦发出淫秽色情言论即可构成犯罪,而超链接具有明确的指向性和引导性,因此也可构成犯罪。在相关的司法解释中可找到具体的定罪标准。

2.2.2.2 编造、故意传播虚假恐怖信息罪

恐怖信息的内涵已从最初的反恐扩展到与社会公共秩序的稳定和安全有关的其他多个方面。一般都具有虚假性、欺骗性、恐怖性和现实可能性。“编造”是指虚构事实、无中生有地凭空捏造,“传播”则无论采用何种方式,只要具备传递散布虚假恐怖信息的行为即可。“严重扰乱社会秩序”的认定则可依据最高人民法院2013年出台的司法解释。笔者也赞同在《刑法修正案(九)(草案)》中增加编造、传播虚假信息罪的相关规定。

2.3 网络失范言论的司法适用疑难解析

2.3.1 现行司法适用规范依据

我国现阶段针对网络失范言论的法律规范主要体现:首先,可以在有关互联安全和信息保护的法律法规中找到一些有关网络失范言论犯罪行为的规定。其次,刑法也针对一些网络犯罪进行了专门性规定,为网络失范言论犯罪的定罪量刑提供了刑法根据。最后,也可以在最高人民法院、最高人民检察院以及其他部门所颁布的司法解释和规范性文件中找到相应的规定对网络失范言论进行刑法规制。

网络失范言论的司法适用疑难主要集中于三个方面:首先,我国网络失范言论相关刑事立法缺乏必要的司法解释或者司法解释本身也存在诸多问题。其次,司法适用过程中易被网络舆论误导或是在基层执法中出现偏差,造成司法适用不当。最后,刑事立法滞后导致司法适用无据。

2.3.2 特殊主体网络失范言论司法适用疑难

2.3.2.1 新闻记者

虽然公民具有获取信息的知情权,但社会分工日益细化,单个主体无法获得全面信息,只能通过新闻记者的报道进行监督。新闻记者作为信息传播的主体,通过及时捕捉信息、仔细进行筛选、客观真实地反映,使民众能够及时准确了解新闻动态,推动了社会的发展进步,但新闻记者也应该严格遵守职业道德、社会公德和良知。在网络言论的发布上,受到其职业特性的影响易被公众所放大,因此,新闻记者应遵循职业伦理要求,承担更多的社会责任,在网络言论的发布上谨言慎行。防止“媒体审判”或“舆论审判”,以免妨碍司法公正。至于该行为是否构成犯罪,笔者认为倘若新闻记者并没有故意虚构事实,只是部分失实,则不必受到刑事追究,倘若其是为了某种个人利益或团体利益,刻意虚构事实、有意引导舆论,则具有刑事违法性,势必要受到刑罚制裁。

2.3.2.2 律师

网络时代的到来使律师获得了更为广阔的言论自由,通过博客、微博、微信等自媒体发布个人诉求,更能直抒胸臆,也便于民众了解案件相关细节,从更深层次上了解案情,满足公民的知情权,同时也能实现舆论监督,对于公正司法具有积极意义。对于律师的网络失范言论缺乏必要规制,将会严重侵害当事人和其他诉讼参与人的利益,导致社会言论秩序混乱,损害司法公正,对司法公信力造成严重损害,甚至通过司法救济等途径也难以消除其危害。我国目前并未有针对律师的庭外涉案言论的完整规范,一些相关标准过于模糊,在司法实践中造成适用困难。笔者认为,刑法应严格适用律师网络言论,防止造成律师对于司法言论的恐慌,以维护社会主义法治文明建设。但同时律师网络言论自由相比普通人而言,是有限度的,律师庭外言论的发布应以不损害国家利益和他人利益为前提,对那些利用专业能力故意误导舆论,以及出于经济利益驱使妨碍司法秩序,以及损害司法公信力的行为绝不姑息,坚决

予以抵制。

2.3.2.3 社会公知

关于社会公知,刑法学界并未有规范定义,笔者认为,这一概念可以泛指活跃于社会公众领域并具有一定社会知名度的主体。社会公知显然已成为一部分网民言论的代言人,针对社会热点问题,社会公知代表民意发声,成为改革的倡导者和拥护者,在舆论监督和维护公共利益等方面都发挥了积极作用,但同时他们也应承担更多的社会责任。随着网络言论的高度自由化,其弱点在网络空间暴露后,易导致出现信任危机:首先是网络言论的不文明化现象严重。其次,发布不实言论或信息,成为“网络大谣”。最后,言行不一,造成社会诚信的缺失。社会公知的言论在网络上的传播速度快、受众广泛、影响深远,因此一旦对外传播,就可视为一种行为,倘若具有危害性,要依法予以制裁。至于是否要用刑法来进行规制,要看其行为是否具有严重的社会危害性,是否构成犯罪。

2.3.3 总体评析

特殊主体由于具备特殊的身份和社会地位,因此具有特定的社会影响力。在发布网络言论时,相比普通人而言,要承担更多的社会责任,不能随心所欲肆意妄为。一旦网络言论失范,就需要法律的干预,至于刑法可否介入,笔者认为应坚持罪责刑相适应原则。犯多大的罪,就要承担多大责任,法院也应给予相应的刑罚。在具体的定罪过程中,首先,要判断其社会危害性,客体可以是个人的人格权、名誉权或隐私权,也可以是社会公共秩序和国家安全。其次,刑事违法性,犯罪必须是触犯刑法的行为。倘若只是一般违法行为则仅需要行政法规进行处罚即可,只有当社会危害性达到了触犯刑法的程度,才能构成犯罪;反之,倘若未达到所要求的程度,就不构成犯罪。再次,应受刑罚处罚性。刑法的可罚性建立在其谦抑性基础上,刑法的谦抑性要求立法者以最小的支出,少用或不用刑罚,而取得最大的社会效益,以便有效预防犯罪和控制犯罪。最后,要考虑法律效果和社会效果的平衡。法律效果是严格适用法律条文所产生的,它符合罪责刑相适应的原则,而社会效果是要实现法律真正的价值,实现法律的秩序、公平和效益等法律目的。

2.4 以域外法外视角的完善建议

2.4.1 网络失范言论刑法规制的比较分析

德国侧重于对公共利益的保护,因而对网络言论进行了严格的立法限制。目前德国对网络言论的立法规范体系主要是以《多元媒体法》为主,辅之以其他相关法律。美国则给予言论自由以高度保障,因此应对网络失范言论,美国主要采用非刑罚化和严厉的民事制裁措施。英国没有针对网络言论制定专门法律,1996年以前,主要依靠援引相关法律法规对网络失范言论进行规制,1996年《3R安全规则》颁布后,英国对网络言论的监控开始以行业自律为主导,政府监管退至次要地位,发挥着辅助作用。新加坡对于网络言论的限制也是比较严格的,主要保护的是国家安全和社会公共利益,通过一些法律法规、分类许可制度以及行业自律等多种手段共同推进网络言论的规范化进程。

经过对比分析可以看出,德国和新加坡对待网络言论的保护属于相对保护,允许相关法律法规对基本人权加以限制,更加看重公共利益的保护。一方面肯定宪法的直接效力,同时另一方面对颁布特别立法使网络失范言论的规制更加具体化。而美国和英国对网络言论的保护属于绝对保护,更加注重个人利益的保护。政府的介入受到严格限制,一旦颁布的法规与言论自由相违背便会被判处违宪。这种严格的违宪审查制度给公民的权利和自由予以最宽泛的保护,导致刑法适用困难重重。笔者认为,我国可以借鉴德国和新加坡的相对保护模式,在保障网络言论自由的同时,也要通过相关部门法来对其加以限制。

2.4.2 网络失范言论刑法规制的完善思考

2.4.2.1 刑事立法层面的思考

2.4.2.1.1 立法完善所应遵循的原则

第一,必要性原则。首先,应对网络失范言论是否属于宪法上的言论自由表达进行规范判断。其次,要注重刑事政策的价值导向,实现网络失范言论刑法规制和网络创新之间的法益平衡。最后,刑法并非规范网络言论的唯一手段,社会道德、行业准则和普通行政处罚也是规范网络失范言论的重要手段,对净化网络风气起到了重要作用。第二,民主性原则。社会主义的本质就是人民当家做主,立法只有体现人民的意志才能真正实现正义的价值。第三,明确性原则。法律条文本身的特点便要求具有明确性,其内容应当为一般人所能理解并预见。

2.4.2.1.2 相关立法建议

在刑法条文的修改上,应同时注重入罪标准和出罪标准。在入罪方面,应当保持明确性和具体性,只有这样才能确保法律适用的可操作性。在出罪方面,应当正确把握言论自由与公民人身权利以及公共秩序之间的平衡,在合理维护后者权益的基础上尽可能地保障公民言论自由。具体就立法完善建议而言,本文主张在《刑法修正案(九)》中增加网络失范言论规制的相应条款,同时防止刑法过度介入。

2.4.2.2 刑事司法层面的思考

对于网络失范言论的刑法规制,本文认为,在刑事司法层面有必要根据社会现实的刑法保护需要,针对现行刑法规范应尽快出台相关司法解释。从司法适用的理念上讲,在刑事司法实践中应当对网络失范言论相关罪名的司法解释进行严格适用。在充分保障公民言论自由的基础上,要对故意制造和传播网络失范言论的主体在必要时进行刑事处罚,尤其要突出对于集团化犯罪的惩罚力度。但同时,对于司法实践中那些事出有因、被害人存在明显过错或是手段过激等的网络失范言论行为,不宜以犯罪论处。

参 考 文 献

- [1] 高铭暄,马克昌. 刑法学[M]. 北京:北京大学出版社,2007:639
- [2] 高铭暄,马克昌. 刑法学[M]. 北京:北京大学出版社,2011:46
- [3] 陈兴良. 刑法的价值构造[M]. 北京:中国政法大学出版社,1998:353
- [4] 陈兴良. 规范刑法学(第二版)下册[M]. 北京:中国人民大学出版社,2008:713
- [5] 张明楷. 刑法学[M]. 北京:法律出版社,2007:652
- [6] 张明楷. 刑法学(第四版)[M]. 北京:法律出版社,2011:828—1030
- [7] 赵秉志. 刑法修正案(七)专题研究[M]. 北京:北京师范大学出版社,2011:153
- [8] 卢建平. 刑事政策与刑法[M]. 北京:北京人民公安大学出版社,2004:3—6
- [9] 王作富. 刑法(第五版)[M]. 北京:中国人民大学出版社,2011:386
- [10] 王作富. 刑法分则实务研究[M]. 北京:中国方正出版社,2003:4—10
- [11] 黄璐,邹军,邹剑. 网络传播法规与道德教程[M]. 上海:复旦大学出版社,2006:32
- [12] 国际人权法教程项目组. 国际人权法教程第二卷(文件集)[M]. 北京:中国政法大学出版社,2002:3—12
- [13] 现代汉语词典(第五版)[M]. 北京:商务印刷馆出版社,2005:53
- [14] 于同志. 网络犯罪(刑事类)[M]. 北京:法律出版社,2008:115
- [15] 李一. 网络行为失范[M]. 北京:社会科学文献出版社,2007:46—48
- [16] 最高人民法院刑事审判一、二、三、四、五庭. 刑事审判参考[M]. 北京:法律出版社,2006(50):39

- [17] 戴长林. 网络犯罪司法实务研究及相关司法解释与适用[M]. 北京:人民法院出版社,2014:135
- [18] 渠敬东. 缺席与断裂—有关失范的社会学研究[M]. 上海:上海人民出版社,1999:29
- [19] 魏永征. 新闻传播法教程[M]. 北京:中国人民大学出版社,2008:44
- [20] 徐久生,庄敬华译. 德国刑法典[M]. 北京:中国法制出版社,2000:56
- [21] 于志刚,于冲. 网络犯罪的裁判经验与学理思辨[M]. 北京:中国法制出版社,2013
- [22] [美] 杰克·道格拉斯等. 越轨社会学[M]. 石家庄:河北人民出版社,1987:53
- [23] [美] 罗杰·费德勒,明安香译. 媒体形态变化:认识新媒体[M]. 北京:华夏出版社,2000:242
- [24] [美] 罗伯特 K 默顿. 当代社会问题[M]. 纽约:哈考特希雷斯约瓦诺维奇有限公司,1991:823
- [25] [日]曾根威彦. 黎宏译. 刑法学基础[M]. 北京:法律出版社,2005:23
- [26] [日]山中敬一. 黎宏译. 刑法各论[M]. 东京:成文堂,2005:666—668
- [27] [日]前田雅英. 刑法各论讲义(第4版)[M]. 东京:东京大学出版会,2007
- [28] 夏征宇. 论宪法视野下的网络表达自由[D]. 上海:华东师范大学,2013
- [29] 陈帅帅. 出售、非法提供公民个人信息罪认定疑难问题研究[D]. 沈阳:沈阳师范大学,2013
- [30] 戴晶莹. 网络失范言论的刑法规制[D]. 上海:华东政法大学,2013
- [31] 邹书剑. 论网络言论自由的法律保障[D]. 沈阳:辽宁大学,2012
- [32] 赵秉志,刘志伟. 论扰乱公共秩序罪的基本问题[J]. 政法论坛,1999(02):69—77
- [33] 陈兴良. 宽严相济刑事政策研究[J]. 法学杂志,2006(01):17—25
- [34] 高涵. 微博转发的从众心理研究[J]. 中州大学学报,2012(03)
- [35] 温辉. 言论自由:概念及边界[J]. 比较法研究,2005(03)
- [36] 张志超. 网络诽谤的构成及认定中的疑难问题[J]. 中国检察官,2008(8)
- [37] 李凤梅. 个人信息安全的刑法保障[J]. 河北法学,2009(12)
- [38] 梁恒. 风险·规制·完善:刑法视域下的个人信息保护[J]. 重庆邮电大学学报,2009(6)
- [39] 贾学胜. 编造、故意传播虚假信息罪之实证解读[J]. 暨南学报(哲学社会科学版),2010(6)
- [40] 李会彬. 网络言论的刑法规制范围—兼评两高《关于办理利用信息网络实施诽谤等刑事案件适用法律若干问题的解释》[J]. 法治研究,2014(3)
- [41] 邓新民. 自媒体:新媒体发展的最新阶段及其特点[J]. 探索,2006(2)
- [42] 胡田野. 新媒体时代律师庭外言论的规制[J]. 法学,2014(1)
- [43] 应琛,杨江. 当口水包围法庭[J]. 新民周刊,2013(100):26—30
- [44] 虞继光,李纛. 律师的媒体运用边界及规则探究[J]. 中国广播,2014(6)
- [45] 胡继冬. 论微博大V的形成、影响与社会管理[J]. 思想理论教育,2013,(12)
- [46] 史琳玉. 冀中星维权之旅[J]. 法律与生活,2013(10):12—19
- [47] 孙雪强. “吴虹飞”一案的法律解释与法律适用[J]. 边缘法律论坛,2013(2):21—22
- [48] 邢璐. 德国网络言论自由保护与立法规制及其对我国的启示[J]. 德国研究,2006(3)
- [49] 秦前红、陈道英. 网络言论自由法律界限初探—美国相关经验之评述[J]. 信息安全,2006(5)
- [50] 刘正荣. 互联网立法和内容传播责任[J]. 网络传播,2007(4):24—26
- [51] 杨福忠. 论我国规制网络言论立法之完善[J]. 北方法学,2013(2)
- [52] 李彦峰. 网络环境下言论自由的刑法规制问题研究[J]. 法制与社会,2014(3)
- [53] 王西林. 中国近现代诗词发展[J]. 文化史研究,2004,(5):271—281
- [54] 李长东,毛先渠,吕成义,等. 浅谈通信管道的设计与铺设[J]. 北京邮电大学大学学报(自然科学版),2003,55(2):12—17
- [55] Justice O' Connor concurred in *Gentile v. State Bar of Nevada*, 501 U. S. 1030, 1082, 111 S. Ct. 2720, 115 L. Ed. 2d 888 (1991)
- [56] Janine Robben, *Oregon Media And Courts: A Guideline*, 69 Or. Si. B. Bull. 33(2009)

-
- [57] Susan W. Brenner; Fantasy Crime: The Role of Criminal Law in Virtual Worlds. In: Vanderbilt Journal of Entertainment and Technology Law. Vol 2008, P. 18
- [58] 全国人民代表大会. 刑法修正案(九)(草案)条文. http://www.npc.gov.cn/npc/xinwen/lfgz/flca/2014-11/03/content_1885029.htm
- [59] 中国互联网信息中心. 第35次中国互联网络发展状况统计报告. http://cnnic.cn/gywm/xwzx/rdxw/2015/201502/t20150203_51631.htm
- [60] j 予嘉. 《新快报》记者刘虎微博实名举报高官渎职. <http://media.sohu.com/20130729/n382831755.shtml>
- [61] 杜雯雯. 刘虎律师:警方调查涉及其微博举报多名高官的行为. <http://www.chinanews.com/fz/2013/10-11/5366596.shtml>

The Criminal Regulation on the Anomie of Internet Speech

Zeng Lei

(School of Humanities and Law, North China Univ. of Tech., Beijing, 100144)

Abstract In informational era, illegal and criminal activities caused by the anomie of Internet speech have been increasing. Even though the freedom of expression is protected by the constitution, the exertion of rights should be premised on fulfilling obligations. The authority will be abused and even the human rights and social development will be severely damaged if the anomie of Internet speech is not limited for its serious social harmfulness. Our current criminal policy of combining punishment with leniency requires treating differently when regulating the anomie of Internet speech. For our current criminal law and the matters found in judicial application, the writer suggests they should be improved by adding related terms in amended criminal law, enacting related judicial interpretations and strictly applying the criminal judicial rules.

Key words the anomie of Internet speech; freedom of expression; criminal law boundary; analysis of regulation; suggestions of improvement

V 系统在纹理分类中应用

研究生 王 俊

导师 宋瑞霞 郭芬红

(北方工业大学理学院 北京 100144)

摘 要 为了对尺度和旋转变换下的纹理图像进行正确的分类,将 Radon 变换和 V-系统相结合,提出一种纹理分类的算法。首先利用 Radon 变换将图像的旋转化为平移,再对 Radon 变换后的图像进行 V-变换;利用 V-系统的多小波特性,经过一系列的降采样分解过程得到图像在 V-系统下的各层次能量表达,并将这些能量作为纹理图像的特征描述。由于 V-系统的多小波特性以及 Radon 变换对旋转的消除,使得文中的特征描述在图像的放缩和旋转变换下有较强的鲁棒性。在通用纹理数据库中的纹理分类实验结果,表明了该算法的优越性能。

关 键 词 V-系统; Radon 变换; 多小波; 多分辨; 纹理分类

作者简介 王俊,女,1988 年出生,硕士研究生,主要研究方向为计算机图形学。

E-mail:13720027033@139.com

1 绪论

纹理是由于物体表面的物理属性不同而造成的,不同的物理表面会产生不同的纹理。通过纹理分析进行物体材质的分类是有现实意义的,比如,通过对木材和石材纹理的分析,可以对木材和石材进行分类,从而合理的利用和保护。从计算机视觉上来说,因为纹理在图像中普遍存在,在图像中是一个非常重要的特征,因此纹理的分类可以为图像分类和模式识别提供有力的依据,基于特征描述的纹理分类是图像检索中的一个重要分支,在产品检验、医学图像分析及遥感图像分析等领域均有重要应用。遗憾的是,图像的纹理特征虽然普遍存在却难以准确描述,至今纹理也没有公认的定义,因此纹理分类是一个学术界关注并具有挑战性的问题。

在纹理分类中,特征提取是至关重要的。在过去的几十年中,众多的研究者提出了大量的特征提取方法,这些方法基本上可以分成空域方法和变换域方法两大类。空域方法是针对图像灰度信息进行的研究,1973 年 Haralick 等^[1]提出灰度共生矩阵的特征提取方法,作为纹理分类的空域有效算法,为后来人们对纹理的分类提供了理论支持和技术积累。1984 年,Pentland^[2]指出分形模型非常适合于描述纹理图像,之后 Sarkar 等^[3]提出的差分计盒算法,是一种简单快速精度高的分形维数计算方法,并将之应用到纹理分类中,关于分形理论在纹理分类中的研究至今仍是热点^[4]。近年来,引人瞩目的是 Ojala 等^[5]于 2002 年提出的局部二进制模式,该方法分析计算复杂度小,具有多尺度特性和旋转不变特性,在纹理检索领域得到应用和发展,成为空域上纹理分类算法的研究焦点之一^[6]。变换域方法是利用变换系数来进行图像纹理特征分析,较空域方法有更多的优势。最典型的是小波的方法,90 年代以后,小波理论的出现为时频多尺度分析提供了一个更为精确而统一的框架。1989 年,Mallat^[7]首先将小波分析引入纹理分析中之后,随之基于小波的纹理分析方法不断涌现出来。如 Chang 等^[8]提出的基于树结构小波的纹理分类方法,Unser^[9]研究的基于小波框架的纹理分类方法。随着小波理论的

发展,小波在纹理特征提取中的应用也不断发展^[10]。

但是必须注意到,在纹理特征分类中,图像的旋转和缩放会引起纹理图像的变化,但显然这样的变化不应该引起纹理分类的改变。因此在纹理分类中,特别强调尺度和旋转的鲁棒性。小波变换虽然对图像纹理特征的刻画比较准确,但它不能直接解决旋转不变性,因此很多文献针对旋转和尺度的不变进行专题研究,如文献[11]通过对数极坐标与小波变换的结合、文献[10,12]将 Radon 变换与小波变换结合,都较好地解决了纹理图像的旋转不变性,但是这些方法对纹理图像的尺度变化不够鲁棒,在涉及尺度变化的纹理分类中准确率还是不高。

本文的数学工具—— k 次 V-系统^[14],是一类有详细数学表达式的多小波,它由分片 k 次多项式构成,结构简洁使用方便(其中 0 次 V-系统恰是著名的 Haar 小波)。V-系统相对经典小波而言,方便之处在于其详细的数学表达,在使用过程中,只需对处理对象进行相应的数学变换(称之为 V-变换),即可得到由粗到细的多层次特征,本文将 V-系统应用到纹理分类的研究。由于 V-系统的多小波特性的,使得它对尺度变换特别鲁棒,但与大多数小波一样,它对旋转变换也比较敏感。为了解决这一问题,本文在图像作 V-变换之前,先进行 Radon 变换,将图像的旋转化为平移,这样的处理使得算法对旋转有较好的鲁棒性。本文算法较经典小波算法简便,在 Brodatz 纹理数据库的实验结果充分验证了本文算法对尺度和旋转变换的鲁棒性,特别是它在尺度变化的纹理分类中有明显的优势。

2 正文

2.1 V-系统及其多分辨特性

2.2.1 V-系统的数学表达

k 次 V-系统由分段 k 次多项式组成,其详细构造过程可查阅文献[14]。这里仅给出本文用到的 1 次 V-系统的数学表达。

V-系统是分组分类构造的。1 次 V-系统的第一组函数是由 $[0,1]$ 区间上的前两个 Legendre 多项式组成的,它的数学表达式为: $V_1^1(x)=1, V_1^2(x)=\sqrt{3}(1-2x)$;第二组称作 V-系统的两个生成元,它们是 1 次 V-系统中最重要的两个函数,它们生成 V-系统函数中的两个“类”,其数学表达式为:

$$V_2^1(x)=\begin{cases} \sqrt{3}(1-4x), & 0 \leq x < \frac{1}{2}, \\ \sqrt{3}(4x-3), & \frac{1}{2} < x \leq 1. \end{cases} \quad V_2^2(x)=\begin{cases} 1-6x, & 0 \leq x < \frac{1}{2}, \\ 5-6x, & \frac{1}{2} < x \leq 1. \end{cases}$$

上述函数在 $x=\frac{1}{2}$ 处的值是这点左右极限的算术平均值。第 $n(n \geq 3)$ 组是由第二组经压缩、平移、复制得到的,所有的函数按两个生成元分成了两类。其一般表达式为:

$$V_n^{i,j}(x)=\begin{cases} \sqrt{2^n-2}V_2^i\left[2^n-2\left(x-\frac{j-1}{2^n-2}\right)\right], & x \in \left(\frac{j-1}{2^n-2}, \frac{j}{2^n-2}\right) \\ 0 & \text{其他} \end{cases}$$

$$i=1,2, \quad j=1,2,\dots,2^n-2aa, \quad n=3,4,5,\dots.$$

函数系 $\{V_1^1(x), V_1^2(x), V_2^1(x), V_2^2(x), V_n^{i,j}(x), n=3 \cdots; i=1,2; j=1,2,\dots,2^{n-1}\}$ 就是 1 次 V-系统,它是一个由无穷多个分段线性函数构成的 $L^2[0,1]$ 空间上的正交函数系。

2.1.2 V-系统的多分辨特性

令

$$\begin{aligned} V_0 &= \text{span}\{V_1^1(x), V_1^2(x)\}, W_0 = \text{span}\{V_2^1(x), V_2^2(x)\}, W_1 = \text{span}\{V_3^{1,1}(x), V_3^{1,2}(x); V_3^{2,1}(x), V_3^{2,2}(x)\}, \\ W_{n-2} &= \text{span}\{V_n^{1,1}(x), V_n^{1,2}(x), \dots, V_n^{1,2^{n-2}}(x); V_n^{2,1}(x), V_n^{2,2}(x), \dots, V_n^{2,2^{n-2}}(x)\}, n=4,5,\dots \\ V_{j+1} &= V_j \oplus W_j, j=0,1,2,3,\dots \end{aligned}$$

易知有

$$V_j \perp W_j, V_0 \subset V_1 \subset V_2 \subset \dots, \text{且 } \bigcup_{j \geq 0} V_j = L_2[0,1].$$

因此 V-系统有多分辨特性。文献[15]从数学上证明了 V-系统是一类多小波。

2.1.3 V-系统的离散化

一个 2^n 阶的离散 V-矩阵的得到过程是:取 V-系统的前 2^n 个基函数,对第 i 个基函数在 $[0,1]$ 区间均匀地取 2^n 个值,构成一个 2^n 阶方阵的第 i 行,当 i 取遍 $1,2,\dots,2^n$ 时,得到一个 2^n 阶方阵;一般来说这个方阵还不是正交矩阵,那么对它正交化即得到离散正交 V-矩阵。 n 的取值由所处理的信号来决定,对于图像变换取决于图像的大小。

图像的 V-变换:设图像亮度函数为 $f=(f_{ij})_{N \times N}$,则 f 的离散 V-变换为 $F=V_N f V_N^T$,其反变换为 $f=V_N^T F V_N$ 。其中 V_N 是 N 阶 V 矩阵,即一幅数字图像经离散 V-变换,得到频域的一个矩阵,这个矩阵再经相应的 V-逆变换可重构数字图像。

由于 V-变换是正交变换,具有变换后信息集中的特点,一般而言,大多数有用信息将以大系数的形式集中于低频,而噪声及细节信息将以小系数形式存在于高频。

关于 V-系统的更多理论和应用见文献[16]。

2.2 Radon 变换

函数 $f(x,y)$ 的 Radon 变换^[12],指的是沿着直线 l 做线积分,即:

$$R(p,\varphi)\{f(x,y)\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)\delta(p-x\cos\varphi-y\sin\varphi)dx dy,$$

其中 p,φ 分别表示原点到直线 l 的距离和直线 l 与 y 轴的夹角(即直线法线与 x 轴的夹角),它们刻画了直线 l 的位置和方向,即直线 l 的方程为: $p-x\cos\varphi-y\sin\varphi=0$;式中 $\delta(\cdot)$ 表示 Delta 函数,即

$$\delta(t)=\begin{cases} 0, & t \neq 0 \\ \infty, & t = 0 \end{cases}, \quad \int_{-\infty}^{+\infty}\delta(t)dt=1.$$

函数 $f(x,y)$ 的 Radon 变换的极坐标表示为:

$R(p,\varphi)\{f(\rho,\theta)\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\rho,\theta)\delta(p-\rho\cos\theta\cos\varphi-\rho\sin\theta\sin\varphi)|\rho|d\rho d\theta$ 当 $f(x,y)$ 表示一幅图像的灰度时,上式离散化后就是图像的 Radon 变换,即图像的 Radon 变换 $R(p,\varphi)\{f(x,y)\}$ 就是图像沿直线 l 的方向的投影。

当图像 $f(x,y)$ 逆时针旋转 θ_0 角度后得到图像 $f(x,y)$,则 $f(x,y)$ 的 Radon 变换在极坐标下表示为:

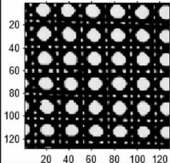
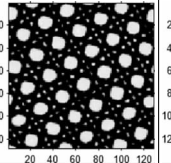
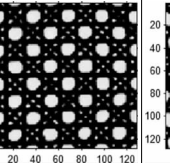
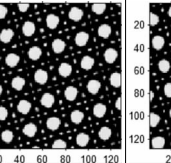
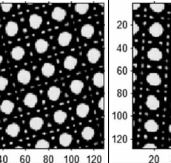
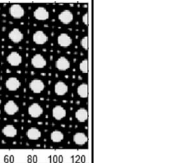
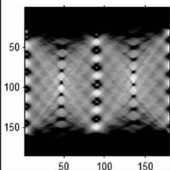
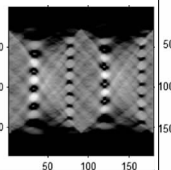
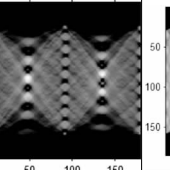
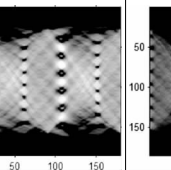
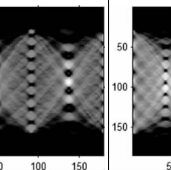
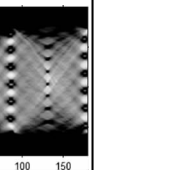
$$\begin{aligned} R(p,\varphi)\{f(\rho,\theta)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\rho,\theta+\alpha)\delta(p-(x\cos\alpha-y\sin\alpha)\cos\varphi-(x\sin\alpha+y\cos\alpha)\sin\varphi)|\rho|d\rho d\theta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\rho,\theta+\alpha)\delta(p-(\rho\cos\theta\cos\alpha-\rho\sin\theta\sin\alpha)\cos\varphi-(\rho\cos\theta\sin\alpha+\rho\sin\theta\cos\alpha)\sin\varphi)|\rho|d\rho d\theta \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(\rho,\theta+\alpha)\delta(p-(\rho\cos(\theta+\alpha))\cos\varphi-(\rho\sin(\theta+\alpha))\sin\varphi)|\rho|d\rho d\theta \\ &= R(p,\varphi)\{f(\rho,\theta+\alpha)\} \end{aligned}$$

即 Radon 变换将图像的旋转转化为变换域图像在角度方向的平移。

表 1 给出了一个纹理图像旋转不同角度后,在 Radon 变换下的结果,显然图像的旋转在 Radon 变换下全部转化成图像的循环平移,Radon 变换的这一性质被广泛地应用于模式识别领域。需要注意的

是，Radon 变换只是对图像重新表达，并不能直接描述对象的特征，需将之与其他特征描述算法相结合，方可解决特征识别中的旋转不变性，如文献[10,12]都是将 Radon 变换与小波变换相结合来解决旋转不变性。

表 1 Radon 变换将图像的旋转化为循环平移

旋转角度	0°	30°	45°	285°	315°	355°
图像						
Radon 变换						

2.3 纹理特征提取及相似度量

V-系统作为一类特殊的正交多小波，虽然可以较准确地描述图像特征，但是它对图像的旋转变换比较敏感，在图像特征识别的时候容易导致误判。为了消除这个影响，本文利用 Radon 变换先将旋转问题转化为平移，然后再经过 V-变换，并做降采样处理，进而提取图像特征。

2.3.1 图像的降采样分解

所谓降采样指的是：分别取图像的奇数行和偶数列，以及偶数行和奇数列，得到两幅分辨率降低 4 倍的图像，这两幅图像称为原图像的降采样图像（见图 1）。

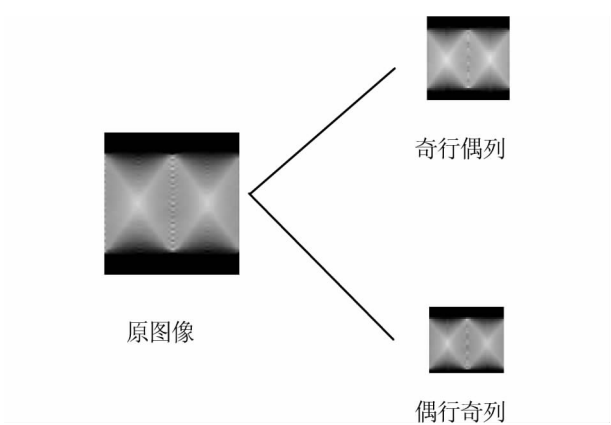


图 1 图像的降采样

一幅图像经 Radon 变换、V-变换之后，再经过降采样过程可以分解为 8 幅分辨率更低的图像，其流程为：

- Step1. 对输入图像 A 进行 Radon 变换，然后进行 V-变换；
- Step2. 将 V-变换域分成一个低频部分、两个中频部分、一个高频部分；

Step3. 将 Step2 中所得矩阵仅保留低频部分,而其余部分设为零,得到一个新矩阵,再对这个新矩阵进行 V-逆变换,得到图像 A1;

Step4. 类似地,将 Step2 中所得矩阵分别仅保留第一中频部分、第二中频部分、高频部分,而相应的其余部分设为零,得到 3 个新矩阵,对这 3 个新矩阵进行 V-逆变换,得到图像 A2—A4;

Step5. 取图像 A1 的两幅降采样图像,记作 A_1^1, A_1^2 ,此即为低频部分的降采样过程。

Step6. 对图像 A2—A4 做同样的降采样,得到中频和高频的 6 幅降采样 $A_2^1, A_2^2, A_3^1, A_3^2, A_4^1, A_4^2$,完成图像的全部降采样。

上述过程将图像 A 分解为 8 幅分辨率降低 4 倍的图像 $A_1^1, A_1^2, A_2^1, A_2^2, A_3^1, A_3^2, A_4^1, A_4^2$,图 2 给出了一个分解示例。若对这样得到的 8 幅图像再分别进行上述 Step1—Step6 的分解,于是原图像就可以分解成一系列的低分辨率图像。上述 V-变换、进而降采样的过程相当于同时完成了小波的高通、低通滤波以及降采样过程,可以使图像的本质特征更加凸显。

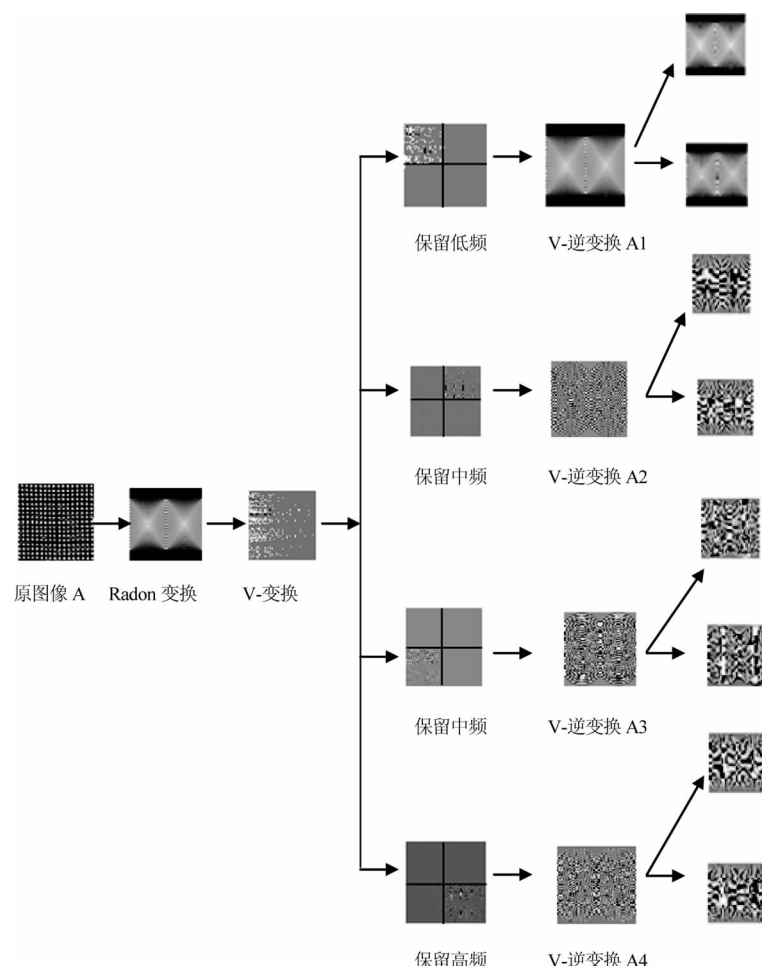


图 2 图像经 Radon 变换和 V-变换后,再降采样分解示例

上述降采样分解,理论上来说可以一直进行到图像只剩下一个像素,但这样的分解显然不必要。故有必要给出恰当的分解终止条件,以减少运算代价。熟知信息熵是对信息量的一种刻画,图像熵的计算公式为: $M = - \sum_{i, x_i \neq 0} x_i \ln x_i$, 其中 x_i 是图像直方图的分布概率,熵越小包含的信息量越大,也即表

明图像表达的内容越多;此外,当图像像素足够少时,信息量也不多了,不必再分解。据此,我们利用熵和图像分辨率来给出图像分解的终止条件。

当图像分解到满足下面两个条件之一时,则终止对图像的分解:

(1)当前图像的熵不大于下次分解所得 8 幅图像熵之均值(说明当前图像的信息量较多,再分解会丢失过多的信息),即 $M_A \leq \frac{1}{8} \sum_{j=1}^4 \sum_{i=1}^2 M_{A_j}^i$;

(2)图像大小为 $N \times N$ (N 是预设的数),这个条件说明图像分解到分辨率足够低时就不再分解。

这样一幅图像可以在计算机程序下自动完成 Radon 变换、V-变换以及恰当的降采样分解。

2.3.2 纹理图像的特征提取及相似度量

一幅纹理图像经过 3.1 节的分解,得到一系列低分辨率的图像。接下来计算每一幅图像的能量: $e = \frac{1}{N^2} \sum_{i,j=0}^{M-1} I_{ij}^2$, 其中 I_{ij} 是图像的灰度, $N \times N$ 是图像的大小。将分解得到的所有图像的能量按降序排列,为了减少运算量、提高效率,我们取前 k 个能量构成图像的特征向量 $\xi = (e_1, e_2, \dots, e_k)$, 通过计算特征向量间的距离、并用最近邻算法就可以进行图像的分类。

本文用到马氏距离和卡方距离,它们的定义分别如下:

马氏距离: $d(x, v_i) = \sqrt{(x - v_i)^T \sum_i^{-1} (x - v_i)}$, 其中 x 是未知图像的特征向量, v_i 是第 i 类图像的均值向量, \sum_i 是第 i 类图像的协方差矩阵;

卡方距离: $d(A, B) = \sqrt{\sum_{i=1}^k \frac{(e_i^A - e_i^B)^2}{e_i^A + e_i^B}}$, 其中 e_i^A, e_i^B 分别表示图像 A, B 的特征向量的第 i 个分量。

经过大量的实验分析,我们发现若要取得最好的分类准确率,对于不同的距离公式,上节终止条件(2)中 N 的取值以及特征向量的维数 k 的取值都不同,实验结果为:马氏距离: $N=32, k=16$;卡方距离: $N=16, k=32$ 。下文中的实验均采用这样的分解数据。

综上,我们给出本文纹理分类算法的流程,如图 3。

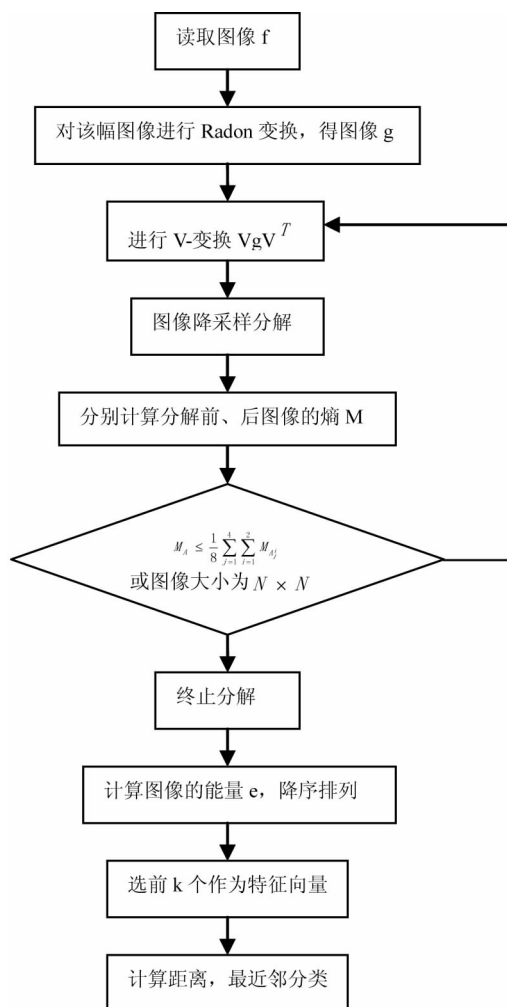


图 3 本文算法流程图

2.4 实验

2.4.1 实验数据库的建立

文献[11,12]都是针对纹理分类中的旋转和尺度问题作了专题研究,本文采用与这两个文献完全相同的实验数据库,实验结果也与这两个文献比较。

文献[11]是从 Brodatz 的数据库中选取如图 4a 所示的 25 幅自然纹理图像,这 25 幅图像按下列操作生成与文献[11]完全相同的 5 个实验数据库:

(1)旋转缩放数据库 A:每幅切割成不重叠的大小为 256×256 的 4 幅纹理图,每一幅图像从 0° 到 345° 每隔 15° 旋转一次,共旋转 24 次;同时,将旋转所得的每一幅图像分别按缩放尺度为 0.6, 0.8, 1, 1.2, 1.4 进行缩放,得到 5 幅缩放图像;共得到 $25 \times 4 \times 24 \times 5 = 12000$ 幅图像。

(2)旋转数据库 B:每幅切割成不重叠的大小为 256×256 的 4 幅纹理图,从 0° 到 355° 每隔 5° 旋转一次,共旋转 72 次,得到 $25 \times 72 \times 4 = 7200$ 幅图像。

(3)缩放数据库 C:每幅切割成不重叠的大小为 256×256 的 4 幅纹理图,从 0.6 到 1.45 每隔 0.05 缩放一次,共计 $25 \times 18 \times 4 = 1800$ 幅图像。

(4)旋转数据库 D:每幅切割成不重叠的大小为 256×256 的 4 幅纹理图,从中选出一幅图像进行

旋转,从 0° 到 355° 每隔 5° 旋转一次,共旋转 72 次,得到 $25 \times 72 = 1800$ 幅图像。

(5) 缩放数据库 E: 每幅切割成不重叠的大小为 256×256 的 4 幅纹理图, 从中选出一幅图像进行缩放, 从 0.6 到 1.45 每隔 0.05 缩放一次, 共计 $25 \times 18 = 450$ 幅图像。

文献[12]也是从 Brodatz 数据库中选的 25 幅自然纹理图(如图 4b, 它与图 4a 中的图像不完全相同), 每幅图像在 0° 到 180° 内旋转 16 次, 旋转角度分别为 $0^\circ, 11^\circ, 22^\circ, 33^\circ, 45^\circ, 56^\circ, 67^\circ, 78^\circ, 90^\circ, 101^\circ, 112^\circ, 123^\circ, 135^\circ, 146^\circ, 157^\circ, 168^\circ$, 共计 $25 \times 16 = 400$ 幅图像, 本文称之为“旋转数据库 F”。

最后将所有图像全部规范成 128×128 大小, 就完成了实验数据库的建立。

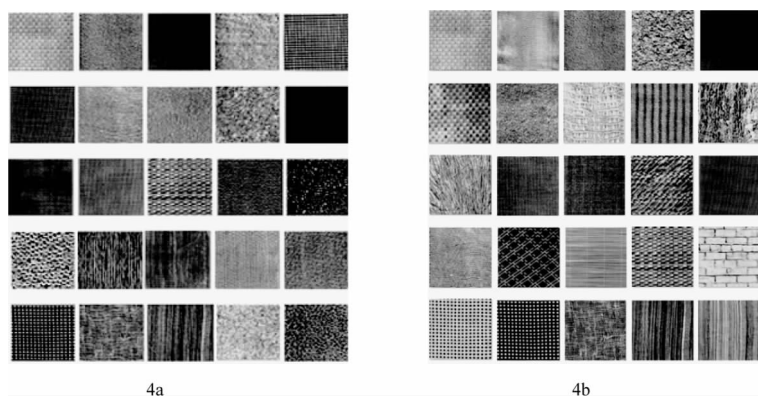


图 4 从 Brodatz 数据库中选取的两组(各 25 幅)自然纹理图

2.4.2 实验结果对比

熟知, 对数极坐标也能将图像的旋转化为循环平移, 文献[11,12]都采用了对数极坐标来处理旋转问题, 此外文献[12]还采用了 Radon 变换与小波变换组合, 文献[12]的实验说明了“对数极坐标+小波”比“Radon 变换+小波”效果更好。为了测试本文算法的有效性, 本文列出“仅用 V—系统”、“对数极坐标+V—系统”、“Radon 变换+V—系统”3 种算法的实验结果, 由于文献[11]采用马氏距离, 文献[12]采用卡方距离, 故本文在与文献[11]比较时用马氏距离, 与文献[12]比较用卡方距离, 详细实验结果分别见表 2 和表 3。

实验结果表明仅用 V—系统准确率不理想, 特别是涉及旋转的纹理图像识别率较低; 将对数极坐标与 V—系统相结合, 在一定程度上提高了准确率, 但在涉及旋转变换的数据库 B、D 中没有超过比较文献[11]中的结果; 在所有实验中, 本文算法即“Radon 变换+V—系统”均取得最好的准确率, 无论是针对旋转变换还是尺度变换, 其分类准确率均达到最高, 说明它对旋转和尺度均有很好的鲁棒性。与文献[11]相比, 本文算法在两个尺度变换数据库“缩放 C”和“缩放 E”上的优势更加明显, 说明本文算法对尺度变换特别鲁棒。从表 2 中我们也注意到, 对缩放数据库 C 和 E, 涉及到 V—系统的 3 种算法(表 2 第 2—4 列)的准确率都较高, 这与 V—系统的结构有关, 因为 V—系统的基函数是通过生成元不同尺度的压缩再复制到各个子区间(见 1.1 节), 且子区间之外的函数值为 0(即函数具有有限支集), 因此 V—系统对尺度变换有一定的适应性。另外, 注意到有的实验数据达到了 100%, 是因为实验数据库比较小, 如果数据库增大, 不能保证达到 100% 的正确率。

表 2 实验结果与文献[11]对比(%) (马氏距离)

数据库	V-系统	对数极坐标+V-系统	Radon 变换+V-系统	文献[11]
旋转+缩放 A	80.93	91.79	94.44	90.80
旋转 B	87.78	93.06	94.99	93.80
缩放 C	92	100	100	88.60
旋转 D	88	98.33	100	100
缩放 E	88	100	100	82.50

表 3 实验结果与文献[12]对比(%) (卡方距离)

	数据库 V-系统	对数极坐标 +V-系统	Radon 变换 +V-系统	文献[12] 对数极坐标 +小波	文献[12] 对数极坐标 +NSCT	文献[12] Radon 变换 +小波	文献[12] Radon 变换 +NSCT
旋转 F	88	96	100	95.5	97.7	58.5	79.4

2.5 结论

Radon 变换可以用来处理旋转问题,但它本身不能直接提取特征,因此经常是将它与其他特征描述方法相结合。本文将 Radon 变换与 V-系统相结合并应用到图像纹理分类,在一定程度上解决了旋转和尺度不变问题。V-系统虽然可以独立提取图像特征,但它对旋转变换不具有不变性,因此本文 Radon 变换在解决旋转不变性中起了重要作用。由于 V-系统对尺度变化特别不敏感,因此本文算法不仅对旋转而且对尺度均有较好的鲁棒性。

V-系统因其有详细的数学表达,而在应用时比通常的小波变换更简洁。V-系统在几何信息特征描述、模式识别、群组对象表达等方面,已经有若干探索例证^[17-19]。这里特别指出,本文基于 V-系统的特征描述与文献[17-19]中的特征描述截然不同。文献[17-19]中是用 V-描述子或 V-矩构成特征向量,而本文是通过图像逐层分解的能量构成特征向量,图像分解过程中采用了降采样方法,并通过图像的熵来决定分解层数,所以本文算法与文献[17-19]有本质不同。

关于 V-系统的应用研究,迄今为止多方面的实验结果表明,V-系统在分离对象的特征表达上有明显优势,因此后续研究应该更加注重 V-系统在群组对象的特征分析方面的研究。

参 考 文 献

- [1] Haralick R M, Shanmugam K, Dinstein I. Textural features for image classification[J]. IEEE Transactions on Systems, Man and Cybernetics, 1973, 3(6): 610-621
- [2] Pentland A P. Fractal based description of natural scenes[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1984, 6(6): 661-674
- [3] Sarkar N, Chaudhuri B B. An efficient differential box-counting approach to compute fractal dimension of image [J]. IEEE Transactions on Systems, Man and Cybernetics, 1994, 24(1): 115-120
- [4] Florindo J B, Bruno O M. Fractal descriptors based on the probability dimension: a texture analysis and classification approach[J]. Pattern Recognition Letters, 2014, 42(1): 107-114

-
- [5] Ojala T, Pietikäinen M, Mäenpää T. Multiresolution gray-scale and rotation invariant texture classification with local binary pattern[J]. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 2002, 24(7): 971—987
 - [6] Yuan F. Rotation and scale invariant local binary pattern based on high order directional derivatives for texture classification[J]. *Digital Signal Processing*, 2014, 26 (1): 142 - 152
 - [7] Mallat S G. A theory for multiresolution signal decomposition: the wavelet representation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1989, 11(7): 674—693
 - [8] Chang T, Kuo C—C. Texture analysis and classification with tree -structured wavelet transform[J]. *IEEE Transactions on Image Processing*, 1993, 2(4) : 429—441
 - [9] Unser M. Texture classification and segmentation using wavelet frames[J]. *IEEE Transactions on Image Processing*, 1995, 4(11): 1549—1560
 - [10] Kourosh J K, Hamid S Z. Rotation-invariant multiresolution texture analysis using radon and wavelet transforms [J]. *IEEE Transactions on Image Processing*, 2005, 14(6): 783—795
 - [11] Pun C—M, Lee M—C. Log-polar wavelet energy signatures for rotation and scale invariant texture classification [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2003, 25(5): 590—603
 - [12] Shao ZhenFeng, Li DeRen, Zhu XianQiang. A multi-scale and multi-orientation image retrieval method based on rotation-invariant texture features[J]. *Science China: Information Sciences*, 2011, 54(4): 732—744
 - [13] Dong Yongsheng, Ma Jinwen. Wavelet-based image texture classification using local energy histograms[J]. *IEEE Signal Processing Letters*, 2011, 18(4): 247—250
 - [14] Song Ruixia, Ma Hui, Wang Tianjun, et al. The complete orthogonal V—system and its applications[J]. *Communication on Pure and Applied Analysis*, 2007, 6(3): 853 - 871
 - [15] Huang Chao, Yang Lihua, Qi Dongxu. A new class of multi-wavelet bases: V—system[J]. *Acta Mathematica Sinica*, 2012, 28(1): 105—120
 - [16] Qi Dongxu, Song Ruixia, Li Jian. Non-continuous orthogonal functions[M]. Beijing: Science Press, 2011; 144—293 (in Chinese)
 - [17] Song Ruixia, Chen Xi, Sun Honglei, et al. A novel algorithm of classification and retrieval for shape group[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2011, 23(12): 1981—1986 (in Chinese)
 - [18] Song Ruixia, Sun Honglei, Wang Xiaochun, et al. Trademark retrieval algorithm based on combination of boundary and region features[J]. *Journal of Software*, 2012, 23(Suppl. (2)): 85—93
 - [19] Song Ruixia, Wang Yena, Wang Xiaochun, et al. Application of V—system on shape similarity retrieval[J]. *Journal of Computer-Aided Design & Computer Graphics*, accepted (in Chinese)

The Application of V system in texture classification

Wang Jun

(School of Economics and Management, North China Univ. of Tech., Beijing, 100144)

Abstract To classify the scaled and rotated texture images correctly, this paper proposes a new algorithm for texture classification by combining Radon transform and the V—system. We firstly use the Radon transform to convert the image rotation into the image translation, and then apply the V—transform on the image obtained after Radon transform. The energies of the image on different levels under the V—system are expressed by performing a series of downsampling process due to the multi-wavelet characteristics of the V—system. These obtained energies are used as the texture feature description. The feature description method in this paper is robust to the image scaling and rotation because of the multi-resolution characteristics of the V—system and elimination of rotation by applying Radon transform. Results of the experiments conducted on the standard texture datasets show that the proposed algorithm provides superior performance.

Key words V—system; radon transform; multi-wavelet; multi-resolution; texture classification

基于光场成像的三维图像特征点提取与匹配

研究生 王晓飞

导师 邹建成

(北方工业大学理学院 北京 100144)

摘 要 图像的特征点提取和匹配是计算机视觉应用中的一个重要组成部分,本文研究的重点是对光场相机曝光得到的多视角平面图像进行特征匹配。实现指定图像对特征匹配的方法有很多种,匹配过程有图像特征点检测、特征描述和图像匹配三个部分。本文围绕这三个方面分别做了研究和探索,详细内容如下:

第一步,图像特征点检测。常用方法有基于图像边缘信息的特征检测方法、基于角点信息的检测方法和各种兴趣算子,但传统的检测方法均存在计算量较大,耗时长的问題,本文针对此问题采用了一种基于图像灰度信息的特征检测方法——FAST 算子,该方法结合决策树理论有效改善了提取图像特征点的速度问题。

第二步,特征描述。特征点描述是根据 BRIEF 描述子对检测到的图像特征点进行一种局部的表达形式。由于 BRIEF 描述子不具备旋转不变性,所以文中先对 FAST 检测算子赋予一个方向,然后对特征描述子进行局部特征描述,生成一个含有方向信息的二进制串数组。

第三步,特征匹配。度量特征相似性的方法有很多,本文通过最邻近搜索的方法寻找欧氏距离最近的特征点对,该方法的计算量较小,实验结果表明,本文的匹配效率较高,速度较快。

通过仿真实验验证,本文基于 ORB 的最邻近搜索匹配算法与经典特征匹配算法相比较,具有较高的匹配准确率,较快的计算速度,这对光场相机处理捕捉的大量阵列图像的计算方面有很大的优势。

关键词 图像匹配; FAST 算子; BRIEF 描述子; 最邻近搜索; 光场成像

作者简介 王晓飞,女,1989 年出生,硕士研究生,主要研究方向为图像处理。

E-mail: wangxiaofei89@sina. cn

1 绪论

随着当今信息技术的快速发展,人们对信息的依赖越来越强,如何获取更多有用的信息变得尤为重要。最直观的获取信息的方法是通过人眼对外部世界进行观察和认知,据统计,通过人眼视觉能够获取外部 80% 的信息^{[1][2]},这足以说明人类的视觉系统能够通过眼睛成像,进而感知生活中的三维物体。另一种获取信息的方法就是由图像间接得到外部信息。摄影可以记录各种各样的图像信息,通过拍摄景物的反射光线使相机内部的感光介质进行曝光的过程,也可以称之为“以光线绘图”,所以一般的成像方式只能获取景物的平面像。若要得到图像中目标景物的三维形态信息,只能采用扫描方式,而这个过程花费的时间较长,不仅影响图像信息获得的实时性,而且影响成像图片的清晰度。

针对传统成像存在的问题,本文引入光场成像^[3]的概念来完成相机中的计算成像,实现了“先拍摄后对焦”的功能,这项功能不仅克服了传统成像得不到三维景物形态信息的局限性,而且也摆脱了对物理器件依赖的束缚性。如图 1-1 是通过光场成像后对同一张图片分别进行重对焦的不同演示。



图 1-1 光场图片的后对焦处理

目前,随着光电技术器件的发展以及光场理论的进一步完善,光场成像技术正逐步应用到航空拍摄、安全监视、立体显示等领域。

一次拍照便能得到物体的三维光场信息,这是光场成像的最大特点,但在成像过程中记录的仅是曝光景物光线的位置和方向信息,并不能直观看到所见物体的立体像,所以我们需要先处理光线信息,而后得到清晰的“所见”三维图。这个处理方法主要包括对图像进行预处理,计算图像中物体的景深信息,提取图像的特征点并匹配图像,拼接图像并显示。其中图像的特征点提取和匹配是计算机视觉应用中的组成部分,在图像三维重构中有着不可或缺的地位,本文工作的主要内容就是基于光场成像的三维图像进行特征点提取与匹配,研究已有的特征匹配方法,分析其优缺点,并在已有算法的基础上进行完善,从而实现在特征提取、匹配上速度和正确率的大幅度提升。

2 基于微透镜阵列的光场成像研究

2.1 光场信息采集

为记录相机中的光场,光场成像中用到了微透镜阵列^[4],放在主镜头和光场传感器之间,每一个微透镜都能记录一小列像元,并且能把光线分离成一个微小的图像,形成入射光的微型图,相机光场的采样是在一次曝光中完成的,微透镜可以被认为输出图像的像素,光电传感器可以被认为输出图像的灰度值。

光学器件成像主要有采集景物的反射光线和分析处理数据。其中光场信息的获取可以分为多相机阵列采集和单相机采集两种方式。多相机阵列是通过多个摄像头同时拍摄一个目标物体,由于相机所处的位置不同,所以采样光场的方向信息和位置信息也不同,再经过相机中的探测器合成成像。单相机采集光场的方法是引入一个光学器件,改变相机内部四维光场分布成像。对记录的光场数据进行处理,用到了射线跟踪技术,这个方法是假设一个相机的配置是最优的,并且光的传播路径是通过光学器件到成像平面,最后这些光线在虚拟镜像上生成我们想要的图片。这种追踪技术的框架也可用在非收敛非期望的射线中,但前提是要假设射线是以固定的方式聚焦,这样才能完成成像计算。

2.2 光场三维图像匹配处理

目标景物周围的光线投射到光场相机内部,通过微透镜阵列对物体进行光场数据的采集,再经过成像传感器把光场数据映射到像平面上,后期通过对光场数据的处理,显示出全景光场图片。光场相机成像步骤大致如图 2-1 所示:

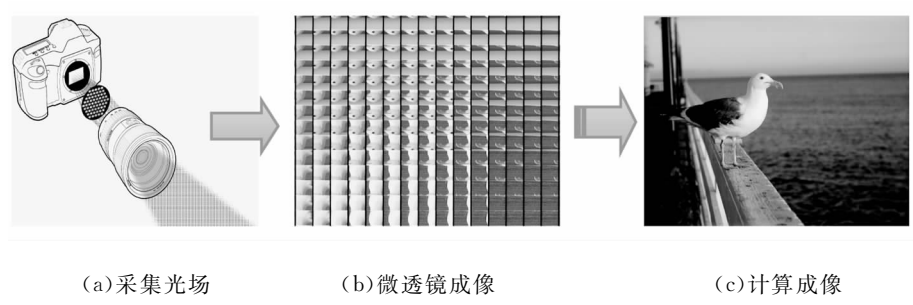


图 2-1 光场相机成像步骤

在微透镜阵列光场相机中,由于微透镜的排列紧凑密集,所以入射到每个透镜的目标景物的光线都可以成一幅完整的且有小角度变化的数字图像,这些数字图像经过图像处理最后能拼接并显示出一幅信息量较大的三维图像,如下图 2-2 是对小角度图像进行特征点提取和匹配,图 2-3 显示出图像匹配后的拼接图像。

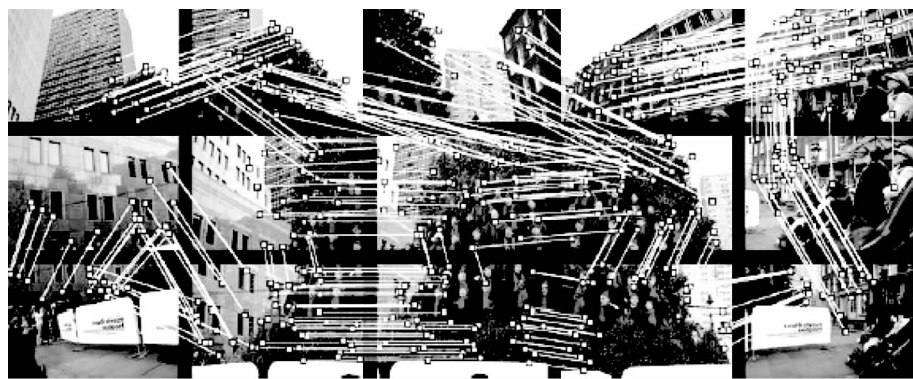


图 2-2 图像特征点提取和匹配



图 2-3 匹配出全景图像并显示

3 基于 ORB 的最邻近搜索匹配算法

3.1 算法简介

ORB(Oriented FAST and Rotated BRIEF)^[5]描述子是由 Ethan Rublee 等人于 2011 年在国际计算机视觉大会上提出的,是 BRIEF^[6]算法的一种改进,具备图像旋转的不变性和抗噪声的特点,并结合目前公认速度较快的 FAST^[7]角点检测算子形成的一个全新的匹配算法。

提取特征点时,ORB 算法在 FAST 角点检测之后赋予特征点一个方向信息,能够保持图像的旋转不变性。匹配特征点时,ORB 算法为了克服 BRIEF 对噪声的敏感性,采用 9×9 的像素块,在像素块中判定灰度值的大小来生成相应的特征描述子,提高了算法对噪声的攻击,且具有旋转不变性。ORB 算法是 FAST 算子和 BRIEF 描述子的结合,融合了两算法计算速度较快的优点。

但 ORB 描述子的缺点也同样存在,它暂时还不能解决图像的尺度不变性问题,并且在匹配时间上还有一定的提升空间。基于 ORB 算法本身的优缺点分析和讨论,本文针对 ORB 描述子对图像匹配的速度问题,提出了一种基于 ORB 的最邻近搜索匹配算法,由 FAST 算子提取出图像的角点信息,然后根据 BRIEF 描述子将图像中感兴趣区域的角点信息表示成二进制串数组,并通过最邻近搜索的方法寻找欧氏距离最近的最佳匹配点。实验结果表明该算法在匹配速度上有明显的提高。

3.2 局部图像特征点检测

本文用 FAST^[8](Features From Accelerated Segment Test)特征检测算子来提取图像的角点信息,这是一种比较快的特征提取方法,该方法主要分为两个步骤:第一步选取出与待检测点相比较的 16 个周边像素点并确定其灰度值;第二步比较检测点与周边像素点的灰度值,根据预设的条件来判定角点。

周边像素值的分布决定了能否提取正确的图像特征点。如果在图像上任取一个像素点 P ,判断点 P 是否为特征点,需要建立一个以 P 点为中心、半径 $r=3$ 的圆作为一个模板图,此模板图是一个离散的圆,如图 3-1 所示,这个离散圆的生成需要用到计算机图形学中的 Bresenham 画圆^[9]算法,并经过多次迭代并遍历 16 个像素点而成。图 3-1 为生成离散圆的初始位置及生成圆的方向示意图。

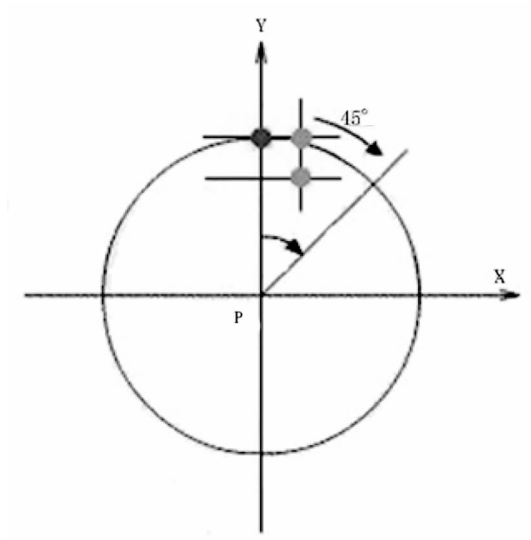


图 3-1 生成圆的初始位置及圆的生成方向

特征点的判定需要在建立的 16 个像素模板图中进行,半径 r 为 3,待检测点为像素点 p ,周边像素值如图中编号,如图 3-2 所示。 p 点的相对位置记为 $p \rightarrow x$,模板上每一个 $x \in \{1, 2, \dots, 16\}$,按照公式 3-1 进行计算:

$$S_{p \rightarrow x} = \begin{cases} d, & I_{p \rightarrow x} \leq I_p - t & (darker) \\ s, & I_p - t < I_{p \rightarrow x} < I_p + t & (similar) \\ b, & I_p + t \leq I_{p \rightarrow x} & (brighter) \end{cases} \quad (3-1)$$

其中 t 为灰度阈值, I_p 是点 p 的灰度值, $I_{p \rightarrow x}$ 为周边像素点的灰度值。最后统计满足公式 $I_{p \rightarrow x} \leq I_p - t$ 和 $I_p + t \leq I_{p \rightarrow x}$ 的像素点的数量 N 。本文用到 FAST-9 的特征检测方法,规定当 $N \geq 9$ 时,既可判定待检测点 p 为特征,图 3-2 为特征点判定的模板示意图。

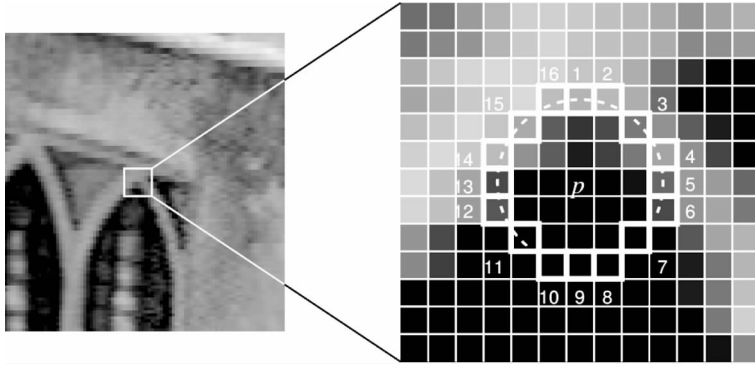


图 3-2 半径为 $r=3$ 的 16 像素模板检测图

角点方向的确定主要分为两部分,首先确定以特征点为中心 7×7 的小像素块的灰度平衡点,也即是质心点;然后计算质心点与特征点之间的偏移量,即是角点的主方向。在圆形邻域半径 r 中, $(x, y) \in [-r, r]$,如下公式:

$$m_{pq} = \sum_{x,y} x^p y^q I(x, y) \quad (3-2)$$

其中 $I(x, y)$ 表示图像 (x, y) 坐标处的灰度值, p 和 q 为阶数。

$$C = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (3-3)$$

C 即是质心点,若设特征点为 P ,则向量 \vec{PC} 就是角点的主方向,用 θ 可表示为:

$$\theta = \arctan\left(\frac{m_{01}}{m_{10}}\right) = \arctan\left(\frac{\sum_{x,y} y I(x, y)}{\sum_{x,y} x I(x, y)}\right) \quad (3-4)$$

其中式子(3-4)中的分母不允许为 0。

3.3 特征描述子

根据 BRIEF 描述子,以特征点为中心,在 31×31 的像素块 P 中,按照 $(X, Y) \sim N(0, \frac{1}{25} S^2)$ 的高斯分布,在 P 中随机产生两组按高斯空间分布的像素点对,比较对应点对的灰度值大小,从而生成一个二进制串描述子,比较公式如下:

定义一个 τ 测试,

$$\tau(P; X, Y) = \begin{cases} 1, & P(X) < P(Y) \\ 0, & P(X) \geq P(Y) \end{cases} \quad (3-5)$$

$$f_n(P) = \sum_{1 \leq i \leq n} 2^{i-1} \tau(P; X_i, Y_i) \quad (3-6)$$

$P(X)$ 是点 X 处的灰度值, $P(Y)$ 是点 Y 处的灰度值, $\tau(P; X, Y)$ 是对检测出的特征点进行的一个二进制串描述, n 为选择的随机点对数, n 可以是 128、256、512 等, 本文选取的 $n=256$ 。

3.4 最邻近特征匹配

最邻近搜索(Nearest Neighbor search, NN)^{[10][11]}的方法能够对检测到的特征点进行快速匹配, 也即是转化成在尺度空间中寻找最近点的优化问题。如果在尺度空间 M 中, 对给定的一个点集 S 和一个目标点 $q, q \in M$, 在 S 中找到距离 q 最近的点 p_i , 通常情况下 M 为欧几里得空间, 距离由欧几里得距离决定。

令 $des(p_i) = (p_{i1}, p_{i2}, \dots, p_{ik})$, $des(q) = (q_1, q_2, \dots, q_k)$, 其中 $k=256$, 则点 $des(p_i)$ 与点 $des(q)$ 之间的欧氏距离可表示为:

$$\begin{aligned} d(des(p_i), des(q)) &:= \sqrt{(p_{i1} - q_1)^2 + (p_{i2} - q_2)^2 + \dots + (p_{ik} - q_k)^2} \\ &= \sqrt{\sum_{k=1}^{256} (p_{ik} - q_k)^2} \end{aligned} \quad (3-7)$$

式中 $des(p_i)$ 是点集 S 中第 p_i 个特征点的特征描述子; $des(q)$ 是目标点 q 的特征描述子; i 为点集 S 中图像特征点的个数; d 表示样本点目标点之间的距离。寻找最短的 d , 也即是找到了与目标点正确匹配的样本点 p_i 。

最邻近搜索算法在图像匹配中的最大缺点是误匹配问题, 所以文中引用了 K -近邻分类算法的思想: 选取样本集中距离目标点最近的 K 个点。本文选取的 K 值为 2, 即找到距离目标点的最短距离 d_{\min} 和次最短距离 d_{\sec} , 为了保证图像匹配的正确性, 规定 $\frac{d_{\min}}{d_{\sec}} \leq \frac{4}{5}$, 且 $d_{\min} \leq 50$ 。

算法步骤如下:

Step1. 初始化最短距离 d_{\min} 和次最短距离 d_{\sec} 为最大值, $d_{\min} = INT_MAX, d_{\sec} = INT_MAX$ 。

Step2. 计算训练样本与每个目标样本点间的距离为 d 。

Step3. 得到目前训练样本中 2 个最邻近目标点的最短距离 d_{\min} 和次最短距离 d_{\sec} 。

Step4. 若 $d < d_{\min}$, 则 d_{\min} 更新为次最短距离, d 更新为最短距离。

若 $d < d_{\sec}$, 则 d 更新为次最短距离。

Step5. 重复步骤 2、3、4, 直到训练样本和目标样本的距离都遍历完, 找到最终的最短距离 d_{\min} 和次最短距离 d_{\sec} 。

Step6. 依据最短距离 d_{\min} 与次最短距离 d_{\sec} 的约束条件, 找到正确的匹配点。

本文基于 ORB 的最邻近搜索匹配算法的流程如下图所示。

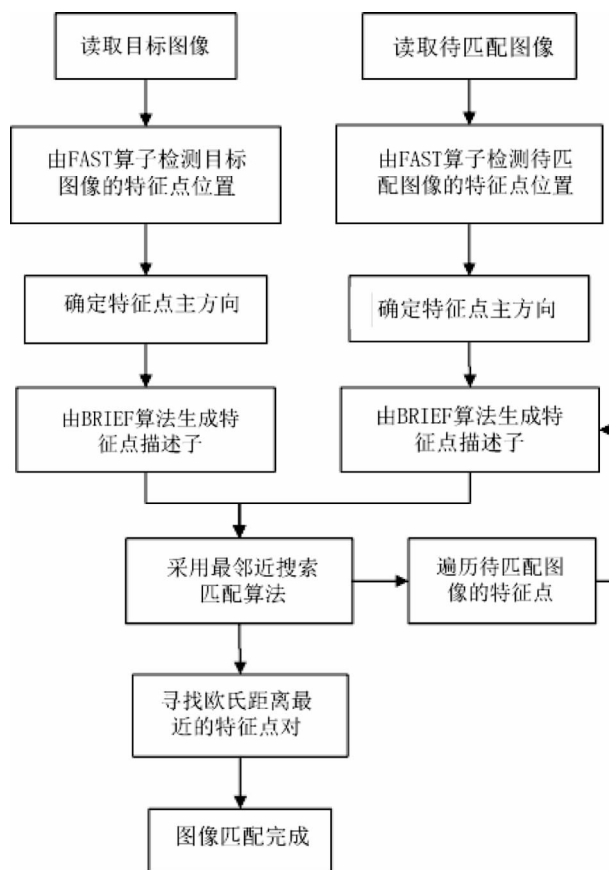


图 3-3 最邻近搜索匹配算法流程图

3.5 实验结果及分析

为了验证本文算法的效果,在 VS2008 环境下,运用 OpenCV 库编写了 C/C++ 程序,使用的计算机配置为 AMD Athlon 64 X2 2.0GHz;2.0GB RAM。本文分别对旋转变化的图像、模糊影响的图像、光照影响的图像和视角变化的图像进行实验分析,使用本文最邻近搜索匹配算法进行实验验证,为了便于观察,如下的匹配效果图只显示了部分正确匹配的特征点对,且选用的实验样例为 12 万像素的图片。

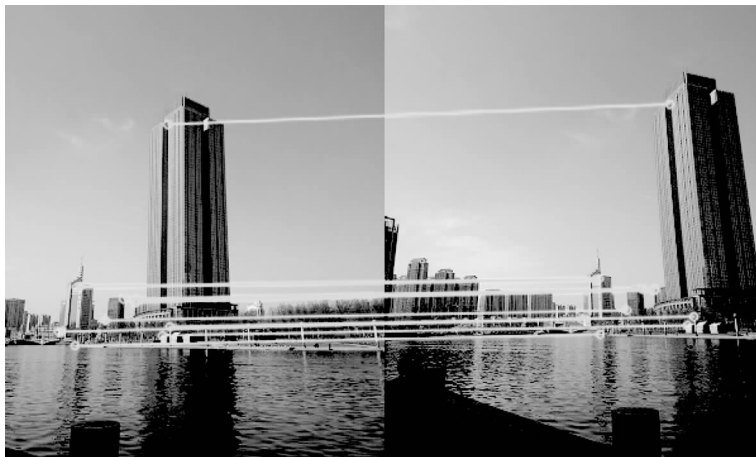


图 3-4 旋转变化的图像匹配效果

(左边图像能提取出 95 个特征点,右边图像能提取出 126 个特征点,能够正确匹配上的点有 65 个,此图只显示了 11 个正确匹配的点)



图 3-5 模糊影响的图像匹配效果

(左边图像能提取出 235 个特征点,右边图像能提取出 323 个特征点,能够正确匹配上的点有 227 个,此图只显示了 13 个正确匹配的点)



图 3-6 光照影响的图像匹配效果

(左边图像能提取出 310 个特征点,右边图像能提取出 411 个特征点,能够正确匹配上的点有 299 个,此图只显示了 11 个正确匹配的点)

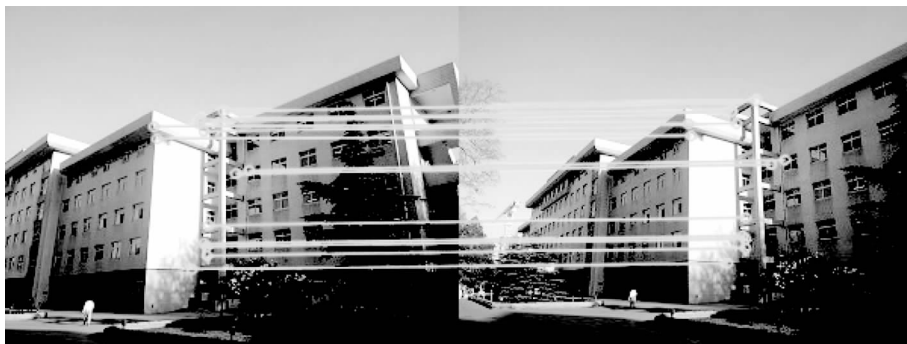


图 3-7 视角变化的图像匹配效果

(左边图像能提取出 196 个特征点,右边图像能提取出 186 个特征点,能够正确匹配上的点有 168 个,此图只显示了 13 个正确匹配的点)

由以上图 3-4 到图 3-7 的 4 幅实验图片的匹配效果验证图可知,本文算法的匹配结果从直观上

看效果比较好,没有误匹配的点对。为了进一步验证本文算法的优势,分别用 SURF 算法、ORB 算法和本文最邻近搜索算法在运算时间和匹配正确率上进行实验分析。验证结果如下表 1 和图 3—8 所示。由表 1 可知,这三种算法在图像特征点提取和匹配上的运算时间相差很大,在特征提取阶段 SURF 算法所用的时间最长,本文算法所需的时间最短;特征匹配阶段 ORB 算法和 SURF 算法对不同场景的图片所需的时间有很大差异,但本文算法要比这两种算法的运算时间有明显优势。图 3—8 是对以上 4 幅测试图像所做的匹配正确率分布柱状图,通过计算正确匹配特征点的数量与提取特征点数量的比值所得。由此图可知本文算法特征点匹配的正确率比 ORB 算法和 SURF 算法的正确率要高。

表 1 不同算法的特征提取与匹配时间对比

算法		SURF	ORB	本文最邻近搜索算法(NN)
图 4—6	特征提取时间(ms)	429	176	11
	特征匹配时间(ms)	16	51	2
图 4—7	特征提取时间(ms)	901	57	14
	特征匹配时间(ms)	58	34	9
图 4—8	特征提取时间(ms)	882	57	16
	特征匹配时间(ms)	59	40	14
图 4—9	特征提取时间(ms)	435	39	9
	特征匹配时间(ms)	27	27	4

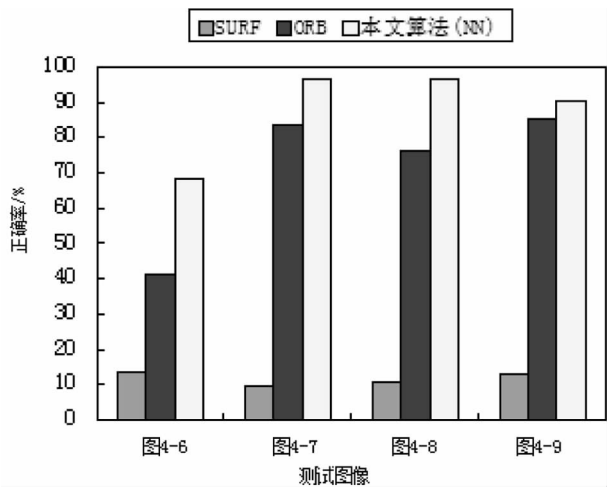


图 3—8 测试图像正确率分布柱状图

保证算法在相同的运算环境中,SURF 算法、ORB 算法和本文最邻近搜索算法的时间结果对比如图 3—9 和图 3—10。图 3—9 和图 3—10 是分别对任意选取的 30 对图像进行的特征点提取时间和特征点匹配时间测试结果折线图,由图 3—9 可知本文算法在特征点提取的时间上要比 ORB 算法略提高了一些,比 SURF 算法要快出近一个数量级。图 3—10 是图像特征点匹配的时间图,在这个阶段本算法在时间上也有一定的优势,本文的最邻近搜索算法要快于 SURF 算法,与 ORB 算法相比,在同等匹配效果下,本文算法的匹配时间明显要快于 ORB 算法。

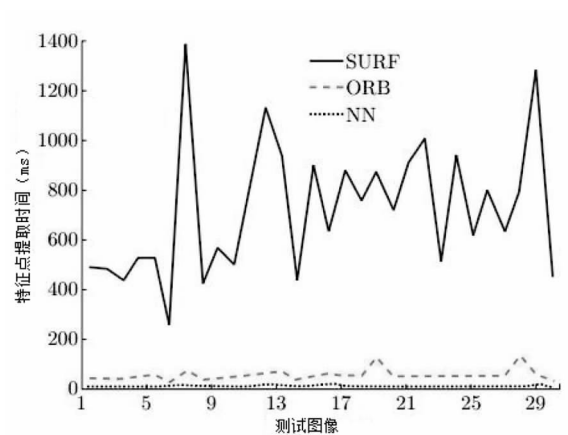


图 3-9 特征点提取时间图

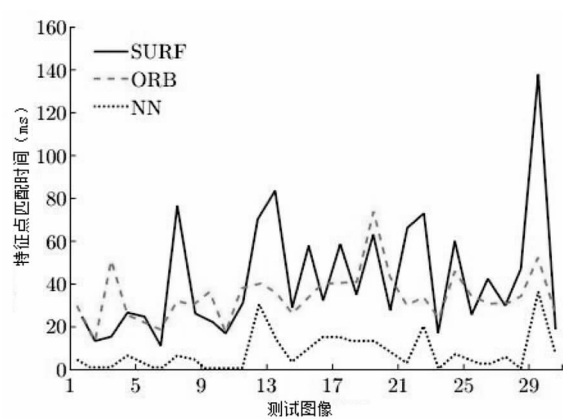


图 3-10 特征点匹配时间图

4 总结

图像特征点提取和匹配是图像三维重构最重要的组成部分,而对光场成像的三维图像进行特征点提取与匹配是一个新颖的研究方向,本文研究的重点是对光场相机曝光得到的多视角平面图像进行特征匹配。近年来对图像匹配的研究比较多,出现很多效率及正确率较高的算法,本文提到的 ORB 算法就是其中一种运算比较快的方法之一,它是 FAST 算子和 BRIEF 描述子的结合,融合了这两种算法的优点。但 ORB 描述子的缺点也同样存在,它暂时还不能解决图像的尺度不变性问题,并且在匹配时间上还有一定的提升空间。所以,对 ORB 算法本身的优缺点进行分析和讨论,针对 ORB 描述子对图像匹配的速度问题,本文提出了一种基于 ORB 的最邻近搜索匹配算法。

该算法在保证匹配效果较好的情况下尽可能的缩短匹配所需的时间。改进后的算法在运算时间和匹配速度上都有很大的改善,主要在于提取角点特征时,只需比较灰度变化较大的区域即可;生成特征描述子时,依据特征点所在的小像素块内的灰度分布即可对特征点进行一个描述;匹配特征点时,采用最邻近搜索方法能够快速找到欧氏距离最近的特征点对。

因此,在当今大数据为背景的时代,对于本文算法匹配时间较快、准确率高的特点,该算法可以应用到检索大量信息匹配的领域,并且对光场相机捕捉到的光束数据进行快速有效的计算。

参 考 文 献

- [1] 张广军. 机器视觉 [M]. 北京:科学出版社,2005
- [2] 马颂德,张正友. 计算机视觉——计算理论与算法基础 [M]. 北京:科学出版社,2003
- [3] Ng R, Levoy M, Bredif M. Light field photography with a hand-held plenoptic camera[R]. Tech Rep CSTR; Stanford Computer Science Tech Report CSTR, 2005
- [4] Ng R. Digital light field photography[D]. USA: STANFORD UNIVERSITY, 2006
- [5] Rublee E, Rabaud V, Konolige K, Bradski G. ORB: an efficient alternative to SIFT or SURF[C] //Proceedings of IEEE International Conference on Computer Vision (ICCV). Los Alamitos: IEEE Computer Society Press, 2011: 2564-2571
- [6] Calonder M, Lepetit V, Strecha C, Fua P. BRIEF: Binary robust independent elementary features[M] //Lecture

Notes in Computer Science. Heidelberg: Springer, 2010, 34(7): 1–28

- [7] Rosten E, Porter R, Drummond T. Faster and better: A machine learning approach to corner detection[J]. IEEE Trans. Pattern Analysis and Machine Intelligence, 2010, 32: 105–119
- [8] Rosten E, Drummond T. Machine learning for high speed corner detection[J]. Proceedings of European Conference on Computer Vision. Springer Berlin Heidelberg, 2006: 32(1): 105–119
- [9] Donald Hearn, M. Pauline Baker. Computer Graphics with OpenGL (3rd Edition). Pearson Education, 2004, ISBN 0-13-015390-7
- [10] Muja M, Lowe D G. Fast approximate nearest neighbors with automatic algorithm configuration[J]. International Conference on Computer Vision Theory and Applications, 2009: 331–340
- [11] Chaudhry S, Xu X, Amato D N. Nearest neighbor search methods[J]. Technical Report Department of Computer Science, Texas A&M University, 2008: 1–8

The feature point extraction and matching based on the three-dimensional imaging of the light field photography

Wang Xiaofei

(School of Economics and Management, North China Univ. of Tech. , Beijing, 100144)

Abstract Image feature extraction and matching is an important component in the applications of computer vision. This paper focuses on the multi-view images matching by the light field camera exposure. So far, there are many ways to achieve the image feature matching, including image feature detection, feature description and matching. This paper has the exploration and research on the three aspects, the details are as follows: The first step, image feature detection. The commonly used methods are the methods based on the image edge information, on corner points information and on various interest operators. But the traditional methods has the problems of the large amount of calculation and time-consuming. In order to solve this problem, this paper uses a feature detection based on the gray information of the image—FAST operator(features from accelerated segment test). This method combined with the decision tree theory is effective to improve the speed of feature extraction of images. The second step, feature description. Description of feature point is a local expression by BRIEF(binary robust independent elementary features). Due to the BRIEF descriptor does not have rotational invariance. So in this paper, we first give a direction, then describe the feature point, and generate a binary string. The third step, image matching. There are many ways to measure the similarity. We use the method of nearest neighbor search to find the feature points by the nearest Euclidean distance. It has simple calculation. The experimental results show that the method has the high efficiency, fast speed. The simulation experiments demonstrated that the nearest neighbor search method for image matching based on ORB has high matching accuracy, fast calculation speed. It also has a great advantage to calculate the large number of image arrays by the light field camera. **Key words** image matching; FAST operator; BRIEF descriptor; the nearest neighbor search; light field photography

基于 V-系统的三维模型数字水印新算法

研究生 徐燕青

导师 宋瑞霞

(北方工业大学理学院 北京 100144)

摘 要 基于一类三角域上定义的正交函数系——V-系统,提出一种在频域嵌入水印的 3D 模型水印新算法。首先针对 3D 模型的顶点信息构造一个“伪模型”,进而对伪模型进行参数化,使之表示为在三角域上定义的一个分片线性函数;然后将这个分片线性函数在 V-系统下进行正交分解,得到伪模型的 V-谱;最后将水印信息通过修改 V-谱嵌入到频域,得到新的 V-谱,利用这些新的 V-谱和 V-系统基函数重构水印伪模型,将水印伪模型的顶点信息替换原 3D 模型的相应顶点,生成最终的水印模型。本文算法水印模型保真度高,且可以在一个模型中植入多个水印依然保持较高的保真度,提取的水印误码率低、相关系数高。理论上证明了本文算法具有相似变换的鲁棒性,通过实验,也表明了本文算法对噪声、剪切和平滑攻击的鲁棒性。

关 键 词 3D 模型;数字水印;V-系统;正交分解;谱

作者简介 徐燕青,男,1989 年出生,硕士研究生,主要研究方向为图形图像处理。

E-mail: xuyanqing1031@126.com

1 绪论

随着多媒体和网络技术的快速发展,3D 模型已经广泛应用于 3D 游戏、3D 动画、3D 影视和艺术品制作等领域,3D 模型的知识产权保护也随之受到关注,其中 3D 数字水印技术作为知识产权保护的重要手段之一,成为学术界的一个研究热点。由于技术上不能简单平移二维图像的水印方案,因此 3D 模型水印远不如图像、音频及视频等数字媒体的水印技术成熟。当然,本质上因为 3D 模型表达形式多样,计算复杂性高,攻击方式多样,所以极具挑战性。

自 1997 年 Ohbuchi 等^[1]第一次提出 3D 模型水印以来,其技术得到迅速发展。根据应用目的不同,3D 模型水印可分为鲁棒水印和脆弱水印两类。脆弱水印对各种攻击极其敏感,任何极其微小的几何或拓扑操作,嵌入其中的水印信息即会消失,因此脆弱水印可以防止非法用户的篡改,适用于保护 3D 模型的完整性。而鲁棒水印在各种攻击下仍能提取水印信息,因此鲁棒水印适用于保护 3D 模型的版权。从水印嵌入算法上来说,可分为空域方法和频域方法,空域方法是直接对模型进行操作,其水印嵌入和提取速度较快,可以嵌入数据量较大的水印,但由于空域水印算法是直接修改网格的几何拓扑连接关系或其它几何不变量去嵌入水印,所以对模型外观影响较大;此外空域算法一般对噪声都比较敏感,对模型的扰动比较大,所以空域水印技术有较多难点要克服,其研究手段也在不断的发展^[2-6]。相对而言,频域水印更加稳定,它是通过将模型变换到频域,在频域进行水印嵌入,然后再重构模型,这样嵌入的水印不会导致模型外观明显的变化,具有较好的保真度,频域水印算法突出的优点是对噪声具有很高的鲁棒性。

常见的频域算法有基于小波的方法^[7-8]、基于球面调和分析的方法^[9-11],利用 Laplace 算子的方

法^[12],利用模型径向基函数构造正交基做谱分析的方法^[13]等。由于小波对信号处理的多分辨能力,使得基于小波的方法一直是频域方法的关注焦点,并不断地得到发展。另外,当把 3D 模型参数化到球面上,利用球面调和分析方法植入水印,使得球面调和成为频域水印的一个重要理论。总体来说,基于频域算法一般都是通过某种数学变换,将模型变换到频域,通过修改频域信息将水印序列嵌入到模型的多个顶点中,这样有利于提高算法的鲁棒性,但缺点是依赖顶点的排序,而且检测水印大多需要依赖原模型。

无论是小波变换还是其他数学变换,目前在 3D 水印的频域算法中,往往是针对模型的某些属性或对模型做某些降维处理后,来进行小波分解或是其他数学变换的,而不是直接针对模型的三角网格进行数学变换,因为三角域上定义的正交系尚不多见,但是若采用三角域上的正交系直接处理三角网格显然是方便合理的。因此本文基于一类三角域上定义的正交函数系——V-系统,提出一种频域鲁棒水印新算法,该算法是直接针对三角网格将模型参数化,并在 V-系统下正交展开,得到模型的频域谱(这些谱可以精确重构三角网格模型),通过调整频域谱来嵌入水印,实验表明该算法能对抗相似变换、噪声、剪切和平滑的攻击。

2 正文

2.1 三角域上 V-系统简介

k 次 V-系统是一类正交函数系,具有多小波的特征,不仅有 $[0,1]$ 区间上定义的单变量 V-系统^[14],也有三角域上定义的多变量 V-系统^[15]。三角域上定义的 V-系统又有直角坐标和面积坐标两种表示。本文用到三角域上定义的、由面积坐标表示的 1 次 V-系统,它是由无穷多个有详细数学表达式的分片一次多项式构成的正交函数系。它的突出特性表现为:

- (1) 运算简单,基函数都是分片一次多项式;
- (2) 基函数有局部性,函数取值中大量为 0;
- (3) 能精确表达由分片多项式表示的对象,无论是单体对象还是群组对象。

这里第 3 个特性是本文用到的重要特性,将一个 3D 模型参数化为一个分片一次多项式后,就可以用 V-系统精确表达。

三角域 V-系统构造原理虽然简单,但因为分片多项式表达,因此数学表达式比较占篇幅,其详细数学表达可以参考文献^[16]的第六章和第八章,有关三角域上 V-系统的应用探索可参考文献^[17-19]。

为简便,下文的“V-系统”均指“三角域上定义的、面积坐标表示的 1 次 V-系统”。

2.2 三角网络模型的 V-系统分解

本文研究的 3D 模型均为三角网格模型,本节的模型 V-谱分解是水印植入的一个关键步骤。

设三角域 G , 它的顶点的面积坐标为 $(1,0,0)$, $(0,1,0)$, $(0,0,1)$ 。

对一个模型作 V-系统下的正交分解,分两步完成:

2.2.1 模型参数化

(1) 设一个 3D 模型 Ω 包含 N 个三角面片, 每个三角面片 $\Delta P_1 P_2 P_3$ 的顶点坐标为 $P_i = (x_i, y_i, z_i)$, $i=1,2,3$; 设 n 为满足 $4^n \geq N$ 的最小整数, 则对三角域 G 做 n 级剖分(见附录 1), 剖分得到的所有子区域记作 $\{G_i | i=1,2,\dots,4^n\}$, 设 G_i 的三个顶点的面积坐标分别为 (u_i, v_i, w_i) , $i=1,2,3$;

(2) 对于每一个 $i=1,2,\dots,N$, 映射第 i 个三角面片 $\Delta P_1 P_2 P_3$ 到 G , 使得 $\Delta P_1 P_2 P_3$ 的 3 个顶点恰好映射为 G_i 的 3 个顶点。这个过程可由两个线性映射构成: 先映射 $\Delta P_1 P_2 P_3$ 到 G_i , 使得 $\Delta P_1 P_2 P_3$ 的

3 个顶点恰好映射为 G 的 3 个顶点;然后 G 再映射到 G_i ,并使 G 的 3 个顶点分别映射为 G_i 的 3 个顶点。综合这两步,第 i 个三角面片 $\Delta P_1 P_2 P_3$ 到 G_i 的映射关系为:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = X_i U_i \begin{pmatrix} u \\ v \\ w \end{pmatrix} = P_{\Delta_i}(u, v, w), \quad (u, v, w) \in G_i \quad (1)$$

其中

$$X_i = \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix}, \quad U_i = \begin{pmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{pmatrix}$$

这样第 i 个三角面片 $\Delta P_1 P_2 P_3$ 就参数化为 $P_{\Delta_i}(u, v, w), (u, v, w) \in G_i$, 显然 $P_{\Delta_i}(u, v, w)$ 是含有 3 个分量的向量函数,每个分量都是一次多项式。

(3) 按照上述过程,当完成了所有的三角面片的映射后, Ω 就表示为一个定义在三角域 G 上的分片一次多项式:

$$P_{\Delta}(u, v, w) = \begin{cases} P_{\Delta_1}(u, v, w), & (u, v, w) \in G_1 \\ P_{\Delta_2}(u, v, w), & (u, v, w) \in G_2 \\ \vdots \\ P_{\Delta_N}(u, v, w), & (u, v, w) \in G_N \\ 0, & (u, v, w) \in G_i, N < j \leq 4^n \end{cases} \quad (2)$$

则式(2)就是模型 Ω 在 G 上的参数化表达。注意这是一个向量形式表达,每个 $P_{\Delta_i}(u, v, w)$, 都有 3 个分量,每个分量都是一次多项式。即

$$P_{\Delta}(u, v, w) = \{P_{\Delta}^x(u, v, w), P_{\Delta}^y(u, v, w), P_{\Delta}^z(u, v, w)\}^T.$$

上述模型参数化过程示意图如图 1。

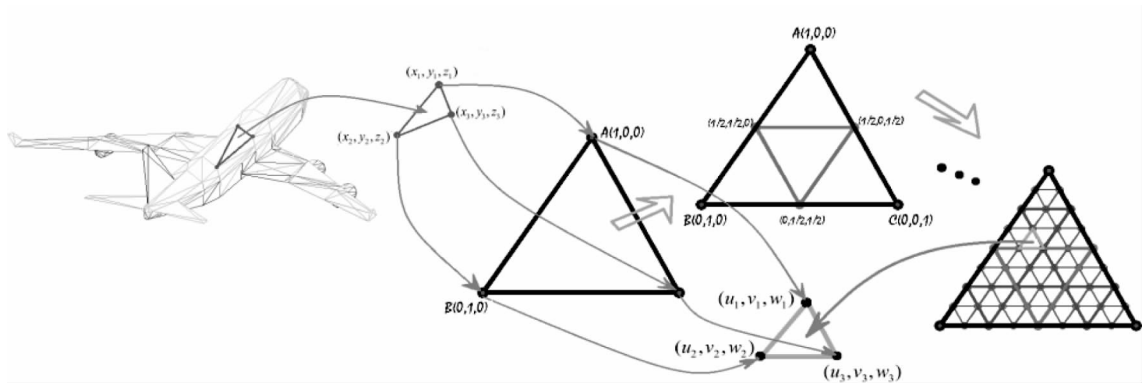


图 1 模型参数化过程示意图

2.2.2 正交分解

设 V-系统的基函数为 $\{V_1(u, v, w), V_2(u, v, w), V_3(u, v, w) \cdots\}$, 则按照上述参数化过程, 含有 N 个面片的模型 Ω 可以分解为:

$$P_{\Delta}(u, v, w) = \{P_{\Delta}^x(u, v, w), P_{\Delta}^y(u, v, w), P_{\Delta}^z(u, v, w)\}^T \quad (3)$$

$$\left\{ \sum_{j=1}^{3 \cdot 4^n} \lambda_x^j V_j(u, v, w), \sum_{j=1}^{3 \cdot 4^n} \lambda_y^j V_j(u, v, w), \sum_{j=1}^{3 \cdot 4^n} \lambda_z^j V_j(u, v, w) \right\}^T$$

其中系数 $\lambda_x^j, \lambda_y^j, \lambda_z^j$ 为:

$$\lambda_x^j = \iint_G P_\Delta^x(u, v, w) V_j(u, v, w) du dv$$

$$\lambda_y^j = \iint_G P_\Delta^y(u, v, w) V_j(u, v, w) du dv$$

$$\lambda_z^j = \iint_G P_\Delta^z(u, v, w) V_j(u, v, w) du dv$$

称 $\{\lambda_x^j, \lambda_y^j, \lambda_z^j\} (j=1, 2, \dots, 3 \cdot 4^n)$ 分别为 X, Y, Z 方向的 V -谱, 式(3)就是模型 Ω 的频域表示, 也叫模型 Ω 在 V -系统下的正交分解。

关于模型的 V -谱, 我们可以证明如下结论(限于篇幅, 此处省略)。

定理 1: (1) 模型经平移变换, 除直流系数(第 0 个 V -谱)之外, 其余 V -谱均不变。

(2) 模型经尺度为 k 的尺度放缩, 则 V -谱也做了相应的尺度放缩, 即

$$(\lambda_x^j, \lambda_y^j, \lambda_z^j) = k(\lambda_x^j, \lambda_y^j, \lambda_z^j)。$$

(3) 模型经旋转矩阵为 R 的旋转变换, 则旋转前后模型的 V -谱的关系为:

$$(\lambda_x^j, \lambda_y^j, \lambda_z^j)^T = R(\lambda_x^j, \lambda_y^j, \lambda_z^j)^T。$$

利用 V -谱, 模型可以重构, 重构算法也分为两步:

1) 逆分解

从 $\{\lambda_x^j, \lambda_y^j, \lambda_z^j\} (j=1, 2, \dots, 3 \cdot 4^n)$ 出发, 即有模型 Ω 的频域表示式(3), 进而得到公式(2), 这意味着对每个 G_i , 得到了方程(1)中的 3×3 矩阵 $X_i U_i$;

2) 重构

利用每个子域 G_j 的顶点坐标 $P_i = (u_i, v_i, w_i), i=1, 2, 3$, 根据方程(1)得到面片 $\Delta P_1 P_2 P_3$ 的三个顶点 $P_i = (x_i, y_i, z_i), i=1, 2, 3$, 当所有三角面片的顶点都得到了, 那么就得到了模型 Ω 。

图 2 给出了一个模型的分解及重构例子, 这个模型是由 172 个散乱的三角面片组成, 形状类似球, 每个三角面片都不相连(但有穿插), 见图 2(a)。我们用这个例子说明 V -系统对散乱的群组对象的表达和重构能力, 因为 V -系统中含有间断的基函数, 并且 V -系统有第 1 节所述的性质(3), 所以这个模型尽管很“散乱”, 但与连续网格一样可以在 V -系统下精确表达并重构。图 2(b)为该模型的 V -谱图, (c) 是通过(b)中 V -谱得到的重构图。

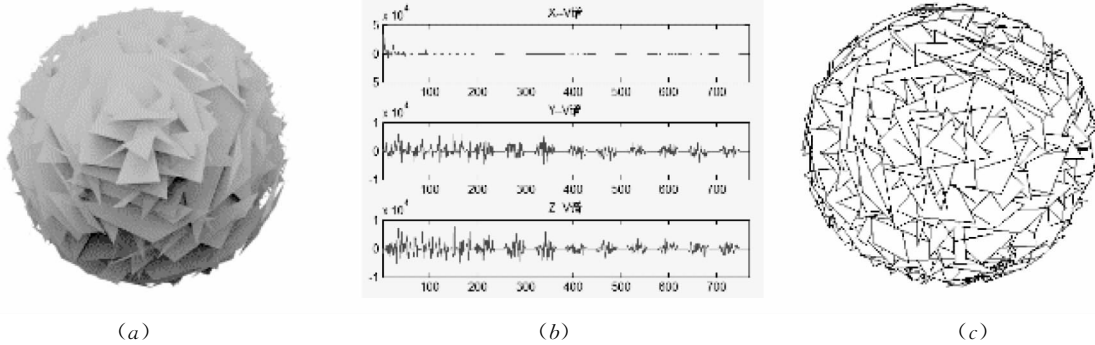


图 2 模型的分解及重构

2.3 水印植入算法

由于 3D 模型的面片数一般都较大, 而水印信息相对较小, 在保证水印提取的质量及水印的鲁棒性的前提下, 为了减少运算量, 我们按如下过程植入水印。

2.3.1 构造伪模型

首先读取原模型的顶点信息和面片索引信息,接着在顶点序列中抽取 $3 \times 4k$ 个顶点(k 的选择与水印信息的长度相关,这些点的抽取方式可以记录下来),这部分点用于构造一个“伪模型”。

“伪模型”的构造:对取出的 $3 \times 4k$ 个顶点,顺次每 3 个构成一个三角面片,这样恰好有 $4k$ 个彼此不相连的三角面片(可能有穿插),这就是我们需要的“伪模型”。这个伪模型除了顶点是原模型的部分顶点外,其余和原模型没有关系,即它的拓扑结构不依赖原模型。伪模型的 $4k$ 个面片恰好可以一一映射到三角域的 k 级剖分域上,由此可以如第 2 节所述对这个伪模型进行 V-系统下的分解与重构。虽然伪模型的每个三角面片都是孤立的,彼此没有关联,但就像图 2 中的例子一样,这不影响它在 V-系统下的分解与重构。

这个伪模型将用于植入水印,同时它也是水印进行非盲检测时的参考模型。

2.3.2 伪模型的 V-谱分解,植入水印

伪模型按照第 2 节所述的方法,进行 V-系统下的分解,得到 XYZ 三个方向的 V 谱,记作 Λ ,设数字水印信息 W 由 0-1 码构成,令

$$new\Lambda = \Lambda + aW$$

其中 a 是一个调控参数,在水印植入过程中可以调节,这样就完成了伪模型谱的修改过程。这里 a 的大小应与模型数据的数量级相关, a 选择太大会对原模型有较大影响,太小则不利于水印的检测,应在一个适度的范围。经过实验发现,若 V 谱中最大的数量级为 10^n ,则 a 取值为 $2 \times 10^{n-4}$ 时效果较好。另外,水印信息可以只植入到 XYZ 三个方向谱中的某一个,也可以分散到三个谱中。

得到新的 V 谱 $new\Lambda$ 后,利用 $new\Lambda$ 按照第 2 节中的重构算法,得到含有水印的伪模型。将水印伪模型的顶点信息读出,并替换原模型的相应顶点,但不改变原模型的边和面的连接关系,这样就生成了与原模型相似的带水印的水印模型,完成了水印植入过程。

这里水印植入过程中,依据模型的顶点构造了一个伪模型,为的是减少计算量(不必对原模型的全部三角网格作 V-谱分解),同时也使得水印模型的保真度更高,特别是提取水印时仅需要参照伪模型,从而提高效率。这种通过伪模型植入水印的算法使得水印信息可以高效率地应用于复杂度高的 3D 模型,并对模型外观影响较小。

2.4 水印提取算法

水印提取过程就是植入过程的逆运算,具体过程为:对待检测模型,首先读取顶点信息,按照生成伪模型时的记录,找到相应的伪模型顶点集,得到 $3 \times 4k$ 个点,只需对该部分进行检测即可。对伪模型中的顶点每三个一组,生成 $4k$ 个散乱三角面片,对这些三角面片进行 V-系统的正交分解,得到 V-谱 Λ' ,同时植入水印时已存有伪模型的 V 谱 Λ ,计算 $W' = (\Lambda' - \Lambda) \cdot a^{-1}$,得到的 W' 即为所提取的水印信息。

2.5 水印评价指标

2.5.1 对水印模型的评价

在频域上嵌入水印的方法就是将嵌入的水印信息分散到 3D 网格的顶点坐标上,本质上修改的就是顶点的坐标,因此我们可以通过计算原网格与水印模型的网格之间对应的顶点的欧氏距离,以此来量化水印对原模型的几何扰动程度。

假设原模型 M 的顶点集为 v_i ,水印模型 M_s 的顶点集为 V'_i ,通过顶点 v_i 与 v'_i 之间的距离定义最小、最大和平均水印扰动距离为^[11]:

$$d_{\min}(M_s, M) = \min(\|v_i - v'_i\| / L), d_{\max}(M_s, M) = \max(\|v_i - v'_i\| / L), d_{\text{aver}}(M_s, M) = \frac{1}{n} \sum_{i=0}^{n-1} (\|v_i - v'_i\| / L)$$

$$v_i - v'_i \| / L),$$

其中 L 表示网格模型包围盒最长对角线的长度, n 是模型的顶点数目。这些距离值越小, 说明模型的保真度越高。

2.5.2 对提取水印的评价

对水印的评价有两个指标——误码率和相关系数。

将提取水印图像与原水印图像比较, 若错误点数为 n , 水印信息包含的总点数为 N , 则对于水印的评价采用误码率来评价, 误码率 $e = n/N$ 。显然 e 的值越小说明提取的水印质量越好, e 的值为 0 时说明提出的水印没有误码、完全正确。

水印相关系数定义为:

$$c = \frac{\sum_{i=1}^N (\omega_i^* - \bar{\omega}^*)(\omega_i - \bar{\omega})}{\sqrt{\sum_{i=1}^N (\omega_i^* - \bar{\omega}^*)^2 \sum_{i=1}^N (\omega_i - \bar{\omega})^2}}$$

其中 ω_i^* 和 ω_i 分别是提取水印信息和原始水印信息, $\bar{\omega}^*$ 和 $\bar{\omega}$ 分别是提取水印信息和原始水印信息的均值。相关系数越接近 1 说明提取的水印质量越好, 相关系数为 1 说明提出的水印与原水印相同。

2.6 算法的鲁棒性

2.6.1 对相似变换的鲁棒性

依定理 1, 模型平移仅直流系数(第 0 个 V-谱)有可能改变, 因此只要水印信息不嵌入到直流系数中, 则平移攻击不会影响水印的提取。

对于尺度和旋转变换, 按照定理 1, 我们可以在对伪模型计算 V-谱时, 根据 V-谱的变化很快判断模型是否做了相应的攻击, 于是在提取水印前找到相关变换参数, 对模型进行相应的逆变换即可。这里为了准确提取变换参数, 只需在水印植入过程中, 保留几个谱不作植入水印时的修改即可。

对尺度攻击, 若找到尺度参数 k , 只要对水印模型作尺度为 $1/k$ 的尺度变换, 就把水印模型还原到原模型大小, 从而顺利提取出水印; 对旋转攻击, 若找到旋转矩阵 R , 则作变换 $M = M'R^{-1}$ 就恢复到旋转之前的位置了, 于是就可以顺利取出水印。

综上, 我们证明了本文算法的水印对平移、尺度、旋转具有鲁棒性。我们也做了相关实验, 在模型分别经过不同程度的平移、放缩、旋转之后, 水印可以完整无误地取出, 这类实验结果恕不赘述。

2.6.2 对噪声、剪切及平滑攻击的鲁棒性

对噪声、剪切及平滑攻击的鲁棒性我们通过实验来说明。

实验中模型选用 Bunny(面片数为 69451)和 Dragon(面片数为 101058)以及 Armadillo(面片数为 345944), 水印图像为: 大小为 99×90 的二值图像, 如图 3 所示。












图 3 水印图像

实验 1 无攻击情形下的模型保真度和水印提取

首先我们实验无攻击情形下,水印模型的保真度和水印提取效果,表 1 是原模型、水印模型、以及提取水印的主观比较和客观评价,由表 1 看出,本文算法的水印模型保真度高,在没有攻击的情形下,可以非常完整地提取水印。

表 1 水印模型和提取水印的主观及客观评价

	原模型	水印的模型	提取的水印	水印评价	水印模型保真度		
					d_aver	d_max	d_min
Bunny				e=0 c=1	6.0942e-05	0.0014	0
Dragon				e=0 c=1	1.2983e-04	0.0014	0
Armadillo				e=0 c=1	4.4151e-05	0.0016	0

实验 2:噪声攻击

实验中噪声是通过对顶点的随机扰动造成的,噪声级是指扰动模型顶点的幅度,比如噪声级为 0.5%,指的是对模型每个顶点做幅度为 0.5%的随机扰动。图 4 给出了水印模型与噪声模型的对比。噪声攻击实验是在水印模型中加入不同等级的噪声,来评价提取水印的效果。我们仅以 Dragon 模型为例进行实验,实验结果见表 2。总体来说,本文算法抗噪性能比较鲁棒,在噪声级达到 0.8%时,依然

能提取比较明确的水印。虽然在噪声攻击时,模型所有的顶点都发生了改变,但并没有严重影响水印的提取质量,这是频域水印的优势。



图4 噪声模型

实验 3:剪切攻击

由于我们的水印是植入在伪模型中,而伪模型的选择相当任意,其面片数可以选择极少,使得够植入水印即可。比如本文水印 0—1 码长度为 8910,伪模型仅需 1024 个面片,涉及 3072 个顶点。如果进一步选择伪模型位于原模型的某个关键的不能被剪切的部位(如龙的头部,因为龙头部一旦剪切,模型将失去使用的意义),则可以抵抗 70%甚至更大比例的剪切攻击,且非常完整地取出的水印;当然,如果伪模型构造的较大,则剪切会对水印提取造成一定的影响,实验结果见表 3。简单说,剪切过程中如果剪切到了伪模型,则提取的水印就会受损,否则水印就可以完整地取出。但是,如果伪模型过小,一旦被全部剪切了,则取不出水印,所以伪模型并不是越小越好。

另一个抗剪切的方案是在模型中构造多个伪模型去植入多个水印,这时对剪切攻击将是非常鲁棒的。植入多个水印,对模型的保真度几乎没有影响,且水印可以完整无误的提取。这类实验是平凡的,此不赘述。

实验 4:平滑攻击

针对 Bunny 模型实验,通过对 Bunny 水印模型进行不同参数下的 Laplacian 平滑(见图 5),再提取水印,并评价其质量,实验结果见表 4。从表 4 和图 5 看到,平滑参数越小,对模型的平滑程度就越低,即模型的改变就小,从而提取的水印质量就好;同一个平滑参数下,平滑多次后模型更平滑,即模型改变也较大,提取的水印质量相对就较差。总体看,本文算法能抵抗一定程度的平滑。



图5 水印模型平滑效果图

表 2 噪声攻击实验结果


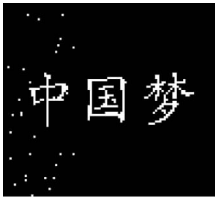
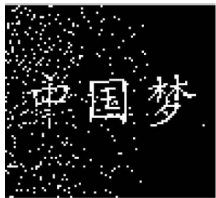

噪声级	0.1%	0.3%	0.5%	0.8%
提取的水印				
误码率	e=0	e= 0.0028	e=0.0379	e=0.1235
相关系数	c=1	c=0.9623	c=0.6739	c=0.4131

表 3 裁剪攻击实验结果










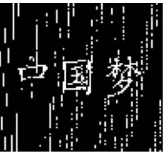
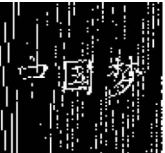
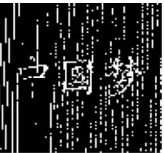
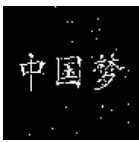
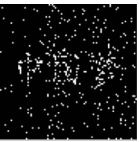



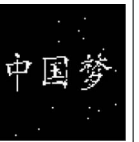
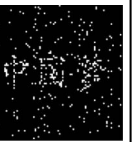
	伪模型较小			伪模型较大		
剪切率	15%	45%	70%	15%	20%	25%
剪切模型						
提取的水印						
水印评价	e=0 c=1	e=0 c=1	e=0 c=1	e =0.0666 c= 0.5385	e = 0.0892 c = 0.4547	e = 0.1097 c = 0.4032

表 4 平滑攻击实验结果

	参数为 1		参数为 0.8		参数 0.5		
	平滑 1 次	平滑 2 次	平滑 1 次	平滑 2 次	平滑 1 次	平滑 2 次	平滑 4 次
提取的水印							
水印评价	e=0.0077 c=0.8964	e=0.0536 c=0.3855	e=0.0017 c=0.9768	e=0.0300 c=0.6030	e=0 c=1	e=0.0046 c=0.9386	e=0.0438 c=0.4574

2.7 结论

本文基于三角域上的正交系 V-系统,提出了一个 3D 模型的水印植入方案,除了理论上保证了具有平移、放缩、旋转的鲁棒性外,从实验结果看,还具有抗噪声污染、抗剪切和抗平滑的鲁棒性。

本文工作新意体现在:(1)水印是通过构造的伪模型来植入的,是一种全新的手段;(2)使用了新的数学工具——三角域上定义的 V-系统,来对伪模型做正交分解,这种正交分解不是对模型做降维处理

再做分解,而是直接对三角网格进行分解,它不同于以往 2D 图像或一维信号的正交分解。

在不降低鲁棒性和保真度的前提下,构造理想的正交系是频域水印算法的关键。目前频域上的水印算法不少,正交系也常见,但三角域上定义的正交系并不多见,而三角域正交系显然是处理三角网格的有利工具。以我们现有的知识,尚未见到将模型在一个三角域上定义的正交系下,依据三角网格做正交展开再植入水印的做法,因此本文是一种新的探索,也是一个挑战。

V-系统作为一种新的数学工具,在本文中起着关键的作用,特别是它的基函数的间断性,使得由一堆散乱三角面片构成的“伪模型”能被精确重构。“伪模型”不仅大大减少了要处理的面片数,而且由于伪模型仅用到原模型的部分顶点信息而不涉及原模型的拓扑信息,因此水印模型的保真度更高;此外伪模型使得水印信息的植入和提取过程都不需要原模型的全部信息,过程非常简单,因此本文水印算法可以高效地处理复杂的 3D 模型。

本文算法除了保持了经典频域算法抗噪声的优点,还对剪切和平滑有较好的鲁棒性。当然它也和很多频域方法一样,不能完全离开原模型取出水印,也还有很多攻击不能抵抗(如模型简化、顶点乱序等),这是将来要继续研究的问题。

参 考 文 献

- [1] Ohbuchi R, Masuda H, Aono M. Watermarking three-dimensional polygonal models[C] // Proceedings of the ACM International Conference on Multimedia97, Seattle, 1997: 261—272
- [2] Praun E, Hoppe H, Finkelstein A. Robust mesh watermarking [C] // Computer Graphics Proceedings, Annual Conference Series, ACM SIGGRAPH, Los Angeles, 1999: 325—334
- [3] Ohbuchi R, Masuda H, Aono M. Watermarking three-dimensional polygonal models through geometric and topological modifications[J]. IEEE Journal on Selected Areas in Communication, 1998, 16 (4): 551—560
- [4] Gao X, Zhang C, Huang Y, et al. A robust high-capacity affine-transformation-invariant scheme for watermarking 3D geometric models [J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMCCAP), 2012, 8(2S): 34
- [5] Su Z, Li W, Kong J, et al. Watermarking 3D CAPD models for topology verification [J]. Computer-Aided Design, 2013, 45(7): 1042—1052
- [6] Pan J, Zheng J, Zhao G. Blind watermarking of NURBS curves and surfaces [J]. Computer-Aided Design, 2013, 45(2): 144—153
- [7] Kanai S, Date H, Kishinami T. Digital watermarking for 3D polygons using multiresolution wavelet decomposition [C] // Proceedings of International Workshop on Geometric Modeling, Tokyo, 1998: 296—307
- [8] Tamane S C, Deshmukh R R. Blind 3D Model Watermarking Based on Multi-resolution Representation and Fuzzy Logic [J]. International Journal of Computer Science & Information Technology, 2012, 4(1)
- [9] Jin J, Dai M, Bao H, et al. Watermarking on 3D mesh based on spherical wavelet transform [J]. Journal of Zhejiang University: Science, 2004, 5 (3): 251—258
- [10] Li L, Zhang D, Pan Z, et al. Watermarking 3D mesh by spherical parameterization [J]. Computers & Graphics, 2004, 28 (6): 981—989
- [11] Wang J R, Feng J Q, Miao Y W. A robust confirmable watermarking algorithm for 3D mesh based on manifold harmonics analysis[J]. The Visual Computer, 28(11): 1049—1062, 2012
- [12] Emad E. A, Ben H, Prabir B. Watermarking 3D models using spectral mesh compression[J]. Signal, Image

- and Video Processing, 2008, 3(4): 375—389
- [13] Wu J H, Leif K. Efficient spectral watermarking of large meshes with orthogonal basis functions[J]. The Visual Computer, 2005, 21(8—10): 848—857
- [14] Song R X, Ma H, Wang T J, et al. The Complete Orthogonal V-system and Its Applications [J], Communication on Pure and Applied Analysis, 6(3), pp. 853 —871, 2007
- [15] Song R X, Wang X C, et al. The Structure of V-system Over Triangulated Domains [J]. Lecture Notes in Computer Science, 4975, pp. 563~569, Springer, 2008
- [16] Qi D X, Song R X, Li J. Non-continuous orthogonal functions[M], Science Press, Beijing, 2011 (in Chinese)
(齐东旭, 宋瑞霞, 李坚. 非连续正交函数[M], 科学出版社, 北京, 2011)
- [17] Li J, Song R X, et al. Orthogonal Reconstruction of 3D Model Base on V-system over Triangular Domain [J], Chinese Journal of Computers, 32(2), pp. 193—202, 2009 (in Chinese)
(李坚, 宋瑞霞, 等. 基于三角域上 V 系统的三维几何模型的正交重构[J], 计算机学报, 2009, 32(2): 193—202)
- [18] Li J, Song R X, et al. The V-system and Expression of Geometric Group Information in Frequency Domain [J], Journal of Software, Vol. 19, Supplement, pp. 41—51, 2008 (in Chinese)
(李坚, 宋瑞霞, 等. V-系统与几何群组信息的频域表达[J], 软件学报, 2008 年增刊: 41—51)
- [19] Song R X, Yao D X, Wang X C, et al. Retrieval Method for 3D Object Group Based on V-system [J], Journal of Advanced Mechanical Design, Systems, and Manufacturing, 6(3), pp. 340—353, 2012

A New Digital Watermark Algorithm for 3D models Based on the V-system

Xu Yanqing

(School of Economics and Management, North China Univ. of Tech., Beijing, 100144)

Abstract Based on a new orthogonal function system, V-system, over triangular domain, this paper proposes a frequency-based watermark algorithm for 3D models. Aiming at the information of the vertex, a “pseudo model” is constructed, and parameterized so that the pseudo model can be expressed as a piecewise linear function over triangular domain. The linear function is then orthogonally decomposed into a V-series, and the V-spectrum is obtained from the coefficients of the V-series. Finally, the watermark is embedded into the frequency domain by modifying the V-spectrum, and a renewed V-spectrum is formed in this way. Using the new V-spectrum and the basis function of the V-system, the pseudo model is reconstructed, and the final watermarked model is built by replacing the vertex of the original 3D model by the vertex information of the watermarked pseudo model. The watermarked model created by the proposed algorithm has high fidelity. Even if multiple watermarks are embedded into a model, the model still maintains high fidelity, and the bit error rate of the extracted watermark is low, the correlation coefficient between the original watermark and the extracted watermark is high. It is theoretically proved that the proposed algorithm is robust against similarity transformations. In addition to this, experiments conducted on watermarked model show that our method is also robust against noise pollution, shearing and smoothing.

Key words 3D model; digital watermark; V-system; orthogonal decomposition; spectrum

非线性 Burgers 方程的高精度数值解法

研究生 范磊

导师 郑权 孙冠颖

(北方工业大学理学院 北京 100144)

摘 要 Burgers 方程是流体力学领域一种重要且基本的抛物型偏微分方程。求解 Burgers 方程的数值方法可以简要归结为有限差分法、有限元法、谱方法和样条函数逼近等几类。本文的主要内容包含两个部分,即分别研究有界和无界区域上的 Burgers 方程的数值解法。

在第一部分,我们研究有界区域上带有非齐次 Dirichlet 边界条件的一维 Burgers 方程的数值解法。利用 Hopf-Cole 变换把原有的非线性问题转化为带有非齐次 Robin 边界条件的线性热传导方程。热传导方程的离散化采用 Crank-Nicolson 差分格式,在两端结合 Crank-Nicolson 格式采用针对 Robin 边界的四阶差分格式。数值方法具有 2 阶收敛性和无条件稳定性,最后通过数值试验检验结论。

在第二部分,我们利用人工边界条件研究无界区域上的一维 Burgers 方程的数值解法。采用 Hopf-Cole 变换,原问题转化为无界区域上的热传导方程问题,并通过引入两个积分形式的人工边界条件把后者简化为在有界区域上的等价问题。针对该等价问题,我们利用降阶方法建立了方程的差分格式和离散化的人工边界条件,进而得到 Burgers 方程的解。采用该方法求解计算区域内的 Burgers 方程被证明并验证具有唯一可解性、稳定性以及具有空间方向上的 2 阶精度和时间方向上的 3/2 阶精度。

关键词 Burgers 方程; Hopf-Cole 变换; Crank-Nicolson 差分格式; 无界区域; 人工边界方法; 有限差分法

作者简介 范磊,男,1989 年出生,硕士研究生,主要研究方向为计算数学。

E-mail: 2311928334@qq.com

1 绪论

Burgers 方程是流体力学领域中一种重要且基本的抛物型偏微分方程[1]。同时,它也是空气动力学和交通流动力学等领域中大量非线性问题的模型方程。它是流体动力学中著名的 Navier-Stokes 方程的简化方程。Burgers 方程首先由 H. Bateman 在 1915 年提出,并由 J. M. Burgers 用于模拟湍流问题。

Burgers 方程在理论和实际应用中地位重要,但由于是非线性的偏微分方程,其精确解不是不存在就是存在却缺乏应用价值,因此发展相应的数值解法就显得十分必要。在众多学者的努力下,有限差分法、有限元法、谱方法以及样条逼近法等数值方法[2—7]相继产生并逐步发展成熟。其中一种有效的方法是利用 Hopf-Cole[8—9]变换把非线性的 Burgers 方程转化为线性的热传导方程,从而降低求解难度、提高计算速度。Kadalbajoo 和 Awasthi[10]采用该变换并结合 Crank-Nicolson 差分格式求解 Burgers 方程,得到了无条件稳定和具有 2 阶精度的结论。关于这类有界区域上的带有齐次边界条件的 Burgers 方程初边值问题,已有的文献资料中包含了大量的研究成果[10—13]。

虽然近来有大量的新方法用于微分方程及其逆问题的求解,有限差分法依旧是一种重要且有效的

方法,并能够在实际应用中与其他方法结合使用[14]。

人工边界方法作为一种有效且成熟的数值方法已被广泛地应用于诸如求解无界区域上的热传导方程、流体力学方程等问题[17,18,28,29,31]。人工边界方法已经被许多学者大量地用于求解无界区域上热传导方程的各类问题,包括二维和三维情形[19–20]。进一步,对于半无界区域 $[-1, \infty) \times [0, \infty)$ 上的热传导方程, Sun 和 Wu[26]首先证明了采用人工边界的差分格式的唯一可解性、无条件稳定性以及具有空间方向上 2 阶和时间方向上 3/2 阶收敛。Wu 和 Zhang[27]也利用人工边界方法得到了关于热传导方程初边值问题的更高阶的差分格式,但仅证明了差分格式的稳定性。

在更进一步的研究中, Han, Wu 和 Xu[22]考虑了无界区域上的非线性 Burgers 方程初边值问题。他们得到了非线性的人工边界条件并建立了非线性的差分格式,没有相应的收敛性分析,只通过数值算例进行了验证。近来, Sun 和 Wu[25]引入了一个函数变换把非线性的 Burgers 方程问题简化为线性方程的初边值问题,建立了一个线性的差分格式,同时证明了差分格式的唯一可解性、无条件稳定性以及具有空间方向上 2 阶和时间方向上 3/2 阶精度。

2 有界区域上 Burgers 方程的有限差分法

2.1 引言

在已有的大量文献中,许多学者仅关注带有齐次边界条件的 Burgers 方程的初边值问题,然而在实际问题中,带有非齐次边界的 Burgers 方程大量存在,因此研究带有更一般的 Dirichlet 边界条件的 Burgers 方程初边值问题很有实际意义。

在本节中,我们求解如下的一维 Burgers 方程:

$$\frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} = \nu \frac{\partial^2 u}{\partial x^2}, (x, t) \in (a, b) \times (0, T], \quad (2.1)$$

初值条件为

$$u(x, 0) = f(x), a < x < b, \quad (2.2)$$

以及非齐次 Dirichlet 边界条件

$$u(a, t) = \alpha(t), u(b, t) = \beta(t), t \in (0, T], \quad (2.3)$$

其中 $\nu = \frac{1}{\text{Re}}$, Re 称为 Reynolds 数, 并且 f, α, β 是给定的足够光滑的函数。

2.2 Hopf-Cole 变换

定理 2.1 (Hopf-Cole) 若 $\varphi(x, t)$ 是如下热传导方程的解

$$\frac{\partial \varphi}{\partial t} = \nu \frac{\partial^2 \varphi}{\partial x^2}, \quad (2.4)$$

则非线性的 Hopf-Cole 变换

$$u = -2\nu \frac{\varphi_x}{\varphi}, \quad (2.5)$$

是 Burgers 方程(2.1)的解。

利用 Hopf-Cole 变换, 初值条件变为

$$\varphi(x, 0) = \exp\left(-\frac{1}{2\nu} \int_a^x u_0(\xi) d\xi\right), a < x < b, \quad (2.6)$$

并且边界条件变为如下 Robin 边界条件:

$$2\nu\varphi_x(a,t)+\alpha(t)\varphi(a,t)=0, 2\nu\varphi_x(b,t)+\beta(t)\varphi(b,t)=0, t>0. \quad (2.7)$$

2.3 热传导方程及其边界条件的离散化

我们用 Crank-Nicolson 差分格式求解热传导方程:

$$\varphi_i^{n+1} - \frac{r}{2}(\varphi_{i+1}^{n+1} - 2\varphi_i^{n+1} + \varphi_{i-1}^{n+1}) = \varphi_i^n + \frac{r}{2}(\varphi_{i+1}^n - 2\varphi_i^n + \varphi_{i-1}^n), i=1\cdots, N-1, \quad (2.8)$$

其中网格比 $r = \frac{\tau}{Reh^2} = \frac{\nu\tau}{h^2}$, 易知 Crank-Nicolson 差分格式具有 $O(\tau^2 + h^2)$ 的精度。

为了离散化(7a), 我们利用 $\varphi_x(0,t) = \frac{\Delta_x\varphi_0}{2h} + O(h^2)$, 其中 $\Delta_x\varphi_i = \varphi_{i+1} - \varphi_{i-1}$ 。则边界条件转化为

$$\frac{2\nu\Delta_x\varphi(x_0,t)}{2h} + \alpha(t)\varphi(x_0,t) = 0. \text{ 接着用 } \frac{\Delta_x}{1 + \frac{1}{6}\delta_x^2} \text{ 代替 } \Delta_x, \text{ 其中 } \delta_x^2\varphi_i = \varphi_{i+1} - 2\varphi_i + \varphi_{i-1}, \text{ 则差分格式变为}$$

如下四阶格式:

$$\frac{\nu(\varphi(x_1,t)) - \varphi(x_{-1},t)}{h(1 + \frac{1}{6}\delta_x^2)} + \alpha(t)\varphi(x_0,t) = 0.$$

我们将 $\varphi_{-1}^n = \frac{(\frac{\nu}{h} + \frac{\alpha^{n+\frac{1}{2}}}{6})\varphi_1^n + \frac{2}{3}\alpha^{n+\frac{1}{2}}\varphi_0^n}{\frac{\nu}{h} - \frac{\alpha^{n+\frac{1}{2}}}{6}}$ 和 $\varphi_{-1}^{n+1} = \frac{(\frac{\nu}{h} + \frac{\alpha^{n+\frac{1}{2}}}{6})\varphi_1^{n+1} + \frac{2}{3}\alpha^n + \frac{1}{2}\varphi_0^{n+1}}{\frac{\nu}{h} - \frac{\alpha^{n+\frac{1}{2}}}{6}}$ 代入 Crank-Ni-

colson 格式并得到

$$\begin{aligned} & (1 + r - \frac{\frac{1}{3}\alpha^{n+\frac{1}{2}}}{\frac{\nu}{h} - \frac{\alpha^{n+\frac{1}{2}}}{6}}r)\varphi_0^{n+1} - (\frac{r}{2} + \frac{\frac{\nu}{h} + \frac{\alpha^{n+\frac{1}{2}}}{6}}{\frac{\nu}{h} - \frac{\alpha^n + \frac{1}{2}}{6}}\frac{r}{2})\varphi_1^{n+1} \\ & = (1 - r - \frac{\frac{1}{3}\alpha^{n+\frac{1}{2}}}{\frac{\nu}{h} - \frac{\alpha^{n+\frac{1}{2}}}{6}}r)\varphi_0^n + (\frac{r}{2} + \frac{\frac{\nu}{h} + \frac{\alpha^{n+\frac{1}{2}}}{6}}{\frac{\nu}{h} - \frac{\alpha^n + \frac{1}{2}}{6}}\frac{r}{2})\varphi_1^n, \end{aligned} \quad (2.9)$$

类似地离散化边界条件(7b), 有

$$\begin{aligned} & (1 + r + \frac{\frac{1}{3}\beta^n + \frac{1}{2}}{\frac{\nu}{h} + \frac{\beta^{n+\frac{1}{2}}}{6}}r)\varphi_N^{n+1} - (\frac{r}{2} + \frac{\frac{\nu}{h} - \frac{\beta^{n+\frac{1}{2}}}{6}}{\frac{\nu}{h} + \frac{\beta^{n+\frac{1}{2}}}{6}}\frac{r}{2})\varphi_{N-1}^{n+1} \\ & = (1 - r - \frac{\frac{1}{3}\beta^n + \frac{1}{2}}{\frac{\nu}{h} + \frac{\beta^{n+\frac{1}{2}}}{6}}r)\varphi_N^n + (\frac{r}{2} + \frac{\frac{\nu}{h} - \frac{\beta^{n+\frac{1}{2}}}{6}}{\frac{\nu}{h} + \frac{\beta^{n+\frac{1}{2}}}{6}}\frac{r}{2})\varphi_{N-1}^n, \end{aligned} \quad (2.10)$$

最后利用离散化的 Hopf-Cole 变换

$$u_i^j = -2\nu \left\{ \frac{\varphi_{i+1}^j - \varphi_{i-1}^j}{2h\varphi_i^j} \right\} = -v \left\{ \frac{\varphi_{i+1}^j - \varphi_{i-1}^j}{h\varphi_i^j} \right\}, \quad (2.11)$$

我们能够得到 Burgers 方程的数值解, 并且易知差分格式(2.8—2.10) 具有时间和空间方向上的二阶精度。

2.4 数值试验

算例 2.1 考虑具有如下精确解的 Burgers 方程:

$$u(x, t) = \frac{0.1e^{-A} + 0.5e^{-B} + e^{-C}}{e^{-A} + e^{-B} + e^{-C}}, A = \frac{0.05}{\nu}(x - 0.5 + 4.95t),$$

$$B = \frac{0.25}{\nu}(x - 0.5 + 0.75t), C = \frac{0.5}{\nu}(x - 0.375).$$

表 2.1 数值解的误差与精确解 ($T=0.1, \nu=0.1, \tau=0.01$)

x	数值解的误差				精确解
	$h=1/10$	$h=1/20$	$h=1/40$	$h=1/80$	
0.1	0.02201986	0.00564128	0.00156442	0.00054623	0.72284902
0.2	0.02085808	0.00527820	0.00139041	0.00041885	0.66530160
0.3	0.01879171	0.00471168	0.00119870	0.00032065	0.60376736
0.4	0.01608664	0.00400578	0.00100416	0.00025483	0.54080409
0.5	0.01323907	0.00328385	0.00081973	0.00020529	0.47917551
0.6	0.01058270	0.00262195	0.00065422	0.00016371	0.42133983
0.7	0.00824508	0.00204590	0.00051235	0.00012996	0.36906159
0.8	0.00616020	0.00153900	0.00039475	0.00010941	0.32326607
0.9	0.00403840	0.00103704	0.00029771	0.00011359	0.28412383

表 2.2 比较不同时间步长下的数值解、精确解及误差 ($\nu=0.1, h=\tau=0.01$)

x	T	数值解	精确解	误差
0.25	0.5	0.77546512	0.77406583	0.00139929
	1.0	0.90009618	0.89819003	0.00190615
	2.0	0.98577150	0.98479730	0.00097420
	3.0	0.99841485	0.99787970	0.00053515
0.50	0.5	0.62550925	0.62456065	0.00094860
	1.0	0.80520920	0.80275849	0.00245071
	2.0	0.96950823	0.96795345	0.00155478
	3.0	0.99632633	0.99566268	0.00066365
0.75	0.5	0.45901235	0.45843125	0.00058110
	1.0	0.65441115	0.65164998	0.00276117
	2.0	0.93356855	0.93077555	0.00279300
	3.0	0.99171892	0.99076624	0.00095268

3 无界区域上 Burgers 方程的人工边界方法

3.1 引言

在实际的科学和工程计算领域,无界区域上偏微分方程的初边值问题大量存在,相较于有界区域

上的问题更具有普遍意义,因而也更具有研究价值。解决无界区域问题的一种重要且有效的方法就是人工边界方法。Han, Wu 和 Xu 考虑了如下的无界区域上的非线性 Burgers 方程初边值问题:

$$u_t - \nu u_{xx} = F(x, t), \forall (x, t) \in R \times (0, T], \quad (3.1)$$

$$u(x, 0) = \varphi(x), -\infty < x < +\infty, \quad (3.2)$$

$$u(x, t) \rightarrow 0, \text{ 当 } |x| \rightarrow +\infty \text{ 时}, \forall t \in [0, T], \quad (3.3)$$

其中 $\nu = \frac{1}{\text{Re}}$, Re 是 Reynolds 数, 给定的函数 F 和 f 充分光滑, 其支集有界, 即 $\text{Supp}\{F(x, t)\} \subset [x_l, x_r] \times [0, T]$ 且 $\text{Supp}\{\varphi(x)\} \subset [x_l, x_r]$ 。

在本文中, 我们研究在 $F=0$ 情况下的问题(3.1)–(3.3), 并在接下来利用 Hopf-Cole 变换把原问题转化为热传导方程的初边值问题。令

$$\Gamma_l = \{(x, t) | x = x_l, 0 \leq t \leq T\},$$

我们得到

$$\Omega_i = \{(x, t) | x_l \leq x \leq x_r, 0 \leq t \leq T\}, \omega(x, 0) = -\int_x^\infty f(y) dy, \text{ 且当 } |x| \rightarrow +\infty \text{ 时}, \omega(x, t) \rightarrow 0.$$

令 $u = \exp(-\omega/2\nu) - 1$, 则原问题(3.1)–(3.3)转化为如下热传导方程的初值问题:

$$\Gamma_l = \{(x, t) | x = x_l, 0 \leq t \leq T\}, \forall (x, t) \in R \times (0, T], \quad (3.4)$$

$$u(x, 0) = \varphi(x) = \exp\left(\frac{1}{2\nu} \int_x^\infty f(y) dy\right) - 1, \quad (3.5)$$

$\Omega_l = \{(x, t) | x \leq x_l, 0 \leq t \leq T\}$, 当 $\Omega_r = \{(x, t) | x \geq x_r, 0 \leq t \leq T\}$ 时, $\Omega_i = \{(x, t) | x_l \leq x \leq x_r, 0 \leq t \leq T\}$ 。

3.2 人工边界条件的推导

首先考虑在无界区域问题(3.4)–(3.6)的解 u 在半无界区域 $[x_r, +\infty) \times [0, T]$ 满足:

$$u_t - \nu u_{xx} = 0, x_r \leq x < +\infty, 0 < t \leq T$$

$$u(x, 0) = 0, x_r \leq x < +\infty,$$

$$u(x, t) \rightarrow 0, \text{ 当 } x \rightarrow +\infty \text{ 时}, 0 \leq t \leq T.$$

当 $u(x_r, t)$ 给定时, 我们得到解 $u(x, t)$:

$$u(x, t) = \frac{x - x_r}{2\sqrt{\pi\nu}} \int_0^t u(x_r, \lambda) (t - \lambda)^{-3/2} e^{-(x - x_r)^2/4(t - \lambda)} d\lambda,$$

令 $\mu = (x - x_r)/2\sqrt{t - \lambda}$, 我们得到:

$$u(x, t) = \frac{2}{\sqrt{\pi\nu}} \int_{(x - x_r)/2\sqrt{t}}^\infty u(x_r, t - \frac{(x - x_r)^2}{4\mu^2}) e^{-\mu^2} d\mu,$$

并且

$$\begin{aligned} \frac{\partial u(x, t)}{\partial x} &= -\frac{2}{\sqrt{\pi\nu}} u(x_r, 0) e^{-(x - x_r)^2/4t} \frac{1}{2\sqrt{t}} \\ &\quad + \frac{2}{\sqrt{\pi\nu}} \int_{(x - x_r)/2\sqrt{t}}^\infty \frac{\partial u}{\partial t}(x_r, t - \frac{(x - x_r)^2}{4\mu^2}) \left(-\frac{2(x - x_r)}{4\mu^2}\right) e^{-\mu^2} d\mu, \end{aligned}$$

变回原变量 λ 可得:

$$\frac{\partial u(x, t)}{\partial x} = -\frac{1}{\sqrt{\pi\nu}} \int_0^t \frac{\partial u(x_r, \lambda)}{\partial \lambda} \frac{1}{\sqrt{t - \lambda}} e^{-(x - x_r)^2/4(t - \lambda)} d\lambda,$$

当 $x \rightarrow +x_r$ 时我们得到 $u_t(x_r, t)$ 和 $u_x(x_r, t)$ 之间的一个关系式:

$$u_x(x_r, t) = -\frac{1}{\sqrt{\pi\nu}} \int_0^t \frac{u_\lambda(x_r, \lambda)}{\partial \lambda} \frac{1}{\sqrt{t - \lambda}} d\lambda.$$

类似地,我们可以得到在 $x=x_l$ 时的人工边界条件。

因此,我们将问题(3.4)–(3.6)简化为如下有界区域上的问题:

$$u_t - \nu u_{xx} = 0, 0 < x \leq T, \quad (3.7)$$

$$u(x, 0) = \varphi(x), x_l \leq x \leq x_r \quad (3.8)$$

$$u_x(x_l, t) = -\frac{1}{\sqrt{\pi\nu}} \int_0^t \frac{u_\lambda(x_l, \lambda)}{\partial \lambda} \frac{1}{\sqrt{t-\lambda}} d\lambda, 0 \leq t < T \quad (3.9)$$

$$u_x(x_r, t) = -\frac{1}{\sqrt{\pi\nu}} \int_0^t \frac{u_\lambda(x_r, \lambda)}{\partial \lambda} \frac{1}{\sqrt{t-\lambda}} d\lambda, 0 \leq t < T. \quad (3.10)$$

3.3 差分格式的推导

引理 3.1 设 $f(t) \in C^2[0, t_n]$, 则有

$$\left| \int_0^{t_n} f'(t) \frac{dt}{\sqrt{t_n-t}} - \sum_{k=1}^n \frac{f(t_k) - f(t_{k-1})}{\tau} \int_{t_{k-1}}^{t_k} \frac{dt}{\sqrt{t_n-t}} \right| \leq \frac{1}{12} (20\sqrt{2} - 23) \max_{0 \leq t \leq t_n} |f(t)| \tau^{\frac{3}{2}}.$$

接着通过引入一个新变量 $v = \frac{\partial u}{\partial x}$ 使得热传导方程(3.7)–(3.10)降阶为等价的一阶微分方程:

$$\frac{\partial u}{\partial t} - \nu \frac{\partial v}{\partial x} = 0, \forall (x, t) \in [x_l, x_r] \times [0, T], \quad (3.11)$$

$$v - \frac{\partial u}{\partial x} = 0, \forall (x, t) \in [x_l, x_r] \times [0, T], \quad (3.12)$$

$$u(x, 0) = \varphi(x), x_l \leq x \leq x_r, \quad (3.13)$$

$$v(x_l, t) = \frac{1}{\sqrt{\pi\nu}} \int_0^t \frac{\partial u(x_l, \lambda)}{\partial \lambda} \frac{1}{\sqrt{t-\lambda}} d\lambda, \quad (3.14)$$

$$v(x_r, t) = -\frac{1}{\sqrt{\pi\nu}} \int_0^t \frac{\partial u(x_r, \lambda)}{\partial \lambda} \frac{1}{\sqrt{t-\lambda}} d\lambda. \quad (3.15)$$

我们对问题(3.11)–(3.15)建立如下差分格式:

$$\delta_i u_i^{n-\frac{1}{2}} - \nu \delta_x v_{i-\frac{1}{2}}^{n-\frac{1}{2}} = 0, 1 \leq i \leq M, n \geq 1, \quad (3.16)$$

$$v_{i-\frac{1}{2}}^{n-\frac{1}{2}} - \delta_x u_i^{n-\frac{1}{2}} = 0, 1 \leq i \leq M, n \geq 1, \quad (3.17)$$

$$u_i^0 = \varphi(x_i), 0 \leq i \leq M, \quad (3.18)$$

$$v_0^{n-\frac{1}{2}} = \frac{2}{\sqrt{\pi\nu}} \left[a_0 u^n - \frac{1}{2} - \sum_{k=1}^{n-1} (a_{n-k-1} - a_{n-k}) u_0^{k-\frac{1}{2}} - a_{n-1} u_0^0 \right], n \geq 1, \quad (3.19)$$

$$v_M^{n-\frac{1}{2}} = -\frac{2}{\sqrt{\pi\nu}} \left[a_0 u_M^{n-\frac{1}{2}} - \sum_{k=1}^{n-1} (a_{n-k-1} - a_{n-k}) u_M^{k-\frac{1}{2}} - a_{n-1} u_M^0 \right], n \geq 1, \quad (3.20)$$

定理 3.1 差分格式(3.16)–(3.20)等价于以下(3.21)–(3.25)。

$$u_i^0 = \varphi(x_i), 0 \leq i \leq M, \quad (3.21)$$

$$\frac{1}{2} (\delta_i u_i^{n-\frac{1}{2}} - \frac{1}{2} + \delta_i u_{i+\frac{1}{2}}^{n-\frac{1}{2}}) - \nu \delta_x^2 u_i^{n-\frac{1}{2}} = 0, 1 \leq i \leq M-1, n \geq 1, \quad (3.22)$$

$$\delta_i u_i^{n-\frac{1}{2}} + \frac{2\nu}{h} \left[\frac{2}{\sqrt{\pi\nu}} (a_0 u_0^{n-\frac{1}{2}} - \sum_{k=1}^{n-1} (a_{n-k-1} - a_{n-k}) u_0^{k-\frac{1}{2}} - a_{n-1} u_0^0) - \delta_x u^n - \frac{1}{2} \right] = 0, \quad (3.23)$$

$$\delta_i u_M^{n-\frac{1}{2}} - \frac{1}{2} + \frac{2\nu}{h} \left[\frac{2}{\sqrt{\pi\nu}} (a_0 u_M^{n-\frac{1}{2}} - \sum_{k=1}^{n-1} (a_{n-k-1} - a_{n-k}) u_M^{k-\frac{1}{2}} - a_{n-1} u_M^0) + \delta_x u_M^{n-\frac{1}{2}} - \frac{1}{2} \right] = 0, \quad (3.24)$$

$$\text{其中 } a_m \equiv \frac{1}{\sqrt{t_{m+1}} + \sqrt{t_m}} = \frac{1}{\sqrt{\tau}(\sqrt{m+1} + \sqrt{m})}, m = 0, 1, 2, \dots. \quad (3.25)$$

3.4 差分格式的理论分析

引理 3.2 对于任意的 $F = \{F_1, F_2, F_3, \dots\}$, 我们有

$$\sum_{l=1}^n [a_0 F_l - \sum_{k=1}^{l-1} (a_{l-k-1} - a_{l-k}) F_k] F_l \geq \frac{1}{2} \sum_{l=1}^n F_l^2, n=1, 2, \dots,$$

其中 a_m 由(3.25)定义。

引理 3.3 若 $\{u_i^n\}$ 是以下问题(3.31)–(3.35)的解

$$\delta_t u_{i-\frac{1}{2}}^{n-\frac{1}{2}} - \nu \delta_x v_{i-\frac{1}{2}}^{n-\frac{1}{2}} = P_{i-\frac{1}{2}}^{n-\frac{1}{2}}, 1 \leq i \leq M, n \geq 1, \quad (3.31)$$

$$v_{i-\frac{1}{2}}^{n-\frac{1}{2}} - \delta_x u_{i-\frac{1}{2}}^{n-\frac{1}{2}} = Q_{i-\frac{1}{2}}^{n-\frac{1}{2}}, 1 \leq i \leq M, n \geq 1, \quad (3.32)$$

$$u_i^0 = \varphi(x_i), 0 \leq i \leq M, \quad (3.33)$$

$$v^n - \frac{1}{2} = \frac{2}{\sqrt{\pi\nu}} \left[a_0 u_0^{n-\frac{1}{2}} - \sum_{k=1}^{n-1} (a_{n-k-1} - a_{n-k}) u^k - \frac{1}{2} - a_{n-1} u_0^0 \right] + S^{n-\frac{1}{2}}, \quad (3.34)$$

$$v^n - \frac{1}{2} = -\frac{2}{\sqrt{\pi\nu}} \left[a_0 u_M^{n-\frac{1}{2}} - \sum_{k=1}^{n-1} (a_{n-k-1} - a_{n-k}) u_M^{k-\frac{1}{2}} - a_{n-1} u_M^0 \right] + T^{n-\frac{1}{2}}, \quad (3.35)$$

其中 $\text{Supp}\{\varphi(x)\} \subset [x_0, x_M]$, 则

$$\begin{aligned} \|u^n\|_A^2 &\leq \exp\left(\frac{2T}{4-\tau}\right) \cdot \left\{ \frac{1}{1-\frac{\tau}{4}} \|u^0\|_A^2 + \frac{\sqrt{\pi\nu}t_n}{2} \tau \sum_{l=1}^n [(T^{l-\frac{1}{2}})^2 + (S^{l-\frac{1}{2}})^2] \right. \\ &\quad \left. + 2\tau \sum_{i=1}^n (\|P^l - \frac{1}{2}\|_A^2 + \|Q^{l-\frac{1}{2}}\|_A^2) \right\}, n=1, 2, \dots. \end{aligned} \quad (3.36)$$

定理 3.2 差分格式(3.21)–(3.25)是唯一可解的。

定理 3.3 令 $\{u_i^n | 0 \leq i \leq M, n \geq 1\}$ 是差分格式(3.21)–(3.25)的解, 则

$$\|u^n\|_A^2 \leq \frac{\exp\left(\frac{2T}{4-\tau}\right)}{1-\frac{\tau}{4}} \|u^0\|_A^2, n=1, 2, \dots. \quad (3.39)$$

定理 3.4 假设问题(3.4)–(3.6)的解为 $u(x, t) \in C_{x,t}^{4,3}(R \times [0, T])$, 记 $\{u_i^n\}$ 是该问题的差分格式(3.21)–(3.25)的解, 并令

$$u_i^n = U_i^n - u_i^n, 0 \leq i \leq M, 0 \leq n \leq [T/\tau],$$

则有

$$\|u^n\|_A^2 \leq \frac{CT}{4-\tau} (\sqrt{\pi\nu T} + 4) \exp\left(\frac{2T}{4-\tau}\right) (\tau^{\frac{3}{2}} + h^2)^2, n=1, 2, \dots, [T/\tau] \quad (3.40)$$

其中 C 是与 τ 和 h 无关的常数。

3.5 数值试验

算例 3.1 我们通过求解初值条件为 $f(x) = \frac{4\nu x}{e^{x^2} + 1}$ 的 Burgers 方程检验所用方法该问题的解析

解为

$$w(x, t) = -2\nu \frac{\frac{1}{2} \frac{\int_{-\infty}^{\infty} \frac{\xi-x}{2\nu t} \exp\left(-\xi^2 - \frac{(x-\xi)^2}{4\nu t}\right) d\xi}{1 + \frac{1}{2} \frac{\int_{-\infty}^{\infty} \exp\left(-\xi^2 - \frac{(x-\xi)^2}{4\nu t}\right) d\xi}{\sqrt{\pi\nu t}}}}{\frac{1}{2} \frac{\int_{-\infty}^{\infty} \frac{\xi-x}{2\nu t} \exp\left(-\xi^2 - \frac{(x-\xi)^2}{4\nu t}\right) d\xi}{1 + \frac{1}{2} \frac{\int_{-\infty}^{\infty} \exp\left(-\xi^2 - \frac{(x-\xi)^2}{4\nu t}\right) d\xi}{\sqrt{\pi\nu t}}}}.$$

表 3.1 关于 τ 的收敛性($\nu=0.5, T=1, h=0.002$ 且 $h=\tau^{3/4}$)

N	M	L^∞ 误差	收敛阶	L^2 误差	收敛阶
10	5000	$3.4349e-04$	—	$5.0342e-04$	—
20	5000	$8.6462e-05$	1.9901	$1.2597e-04$	1.9987
40	5000	$3.6936e-05$	1.2270	$3.2693e-05$	1.9461
80	5000	$3.1874e-05$	0.2126	$1.2036e-05$	1.4416
10	56	$3.6538e-03$	—	$5.3702e-03$	—
20	95	$1.2531e-03$	1.5439	$1.8421e-03$	1.5436
40	159	$4.4110e-04$	1.5063	$6.4865e-04$	1.5058
80	267	$1.5473e-04$	1.5114	$2.2819e-04$	1.5072

表 3.2 关于 h 的收敛性($\nu=0.5, T=1, \tau=0.002$ 且 $\tau=h^{4/3}$)

M	N	L^∞ 误差	收敛阶	L^2 误差	收敛阶
25	200	$1.5826e-02$	—	$2.3957e-02$	—
50	200	$4.1785e-03$	1.9212	$6.2111e-03$	1.9475
100	200	$1.0618e-03$	1.9765	$1.5793e-03$	1.9756
200	200	$2.6775e-04$	1.9876	$3.9703e-04$	1.9920
25	3	$1.7558e-02$	—	$2.8237e-02$	—
50	9	$4.5211e-03$	1.9574	$6.7089e-03$	2.0734
100	22	$1.1260e-03$	2.0055	$1.6559e-03$	2.0185
200	54	$2.7728e-04$	2.0218	$4.0895e-04$	2.0176

4 结论

本文的第一部分针对有界区域上的 Burgers 方程初边值问题,为了克服非线性和非齐次 Dirichlet 边界条件带来的求解困难,我们运用 Hopf-Cole 变换和 Crank-Nicolson 差分格式给出了对它的简便高效的数值解法,同时包含了具有齐次 Dirichlet 边界条件时的情形,证明了差分格式的无条件稳定性和二阶收敛性,并通过数值试验验证了理论。

本文第二部分在 Han, Wu 以及其他学者成果的基础上进行了更深入的研究。本文的引理 3.2 得到了一个比 Wu 和 Sun(2004)以及 Han 和 Wu(2012)的文章中的不等式更强的结果。本文的引理 3.3 的证明运用到了 Gronwall 引理,修正了 Wu 和 Sun(2004)文章中的引理 4 以及 Han 和 Wu(2012)文章中的引理 3.24 的错误,得到了正确的证明。最后,我们证明了差分格式的唯一可解性、无条件稳定性以及具有空间方向上 2 阶和时间方向上 3/2 阶的收敛性,并通过数值试验验证了理论。

参 考 文 献

- [1] J. M. Burgers. A mathematical model illustrating the theory of turbulence[J]. Adv. Appl. Mech., 1984, 1: 171

-
- [2] E. N. Aksan. A numerical solution of Burgers' equation by finite element method constructed on the method of discretization in time[J]. *Appl. Math. Comput.*, 2005, 170: 895–904
 - [3] R. C. Mittal, P. Singhal. Numerical solution of Burgers equation[J]. *Commun. Numer. Methods Eng.*, 1993, 9: 397–406
 - [4] S. Kutluay, A. Esen, I. Dag. Numerical solutions of the Burgers' equation by the least-squares quadratic B-spline finite element method[J]. *J. Comput. Appl. Math.*, 2004, 167: 21–33
 - [5] D. J. Evans, A. R. Abdullah. The group explicit methods for the Burgers' equation[J]. *Computing.*, 1984, 32: 239–252
 - [6] A. G. Bratsos. A fourth-order numerical scheme for solving the modified Burgers equation[J]. *Comput. Math. Comput.*, 2010, 60: 1393–1400
 - [7] P. -G. Zhang, J. -P. Wang. A predictor-corrector compact finite difference scheme for Burgers' equation[J]. *Appl. Math. Comput.*, 2012, 219: 892–898
 - [8] E. Hopf. The partial differential equation $u_t + uu_x = \nu u_{xx}$ [J]. *Commun. Pure Appl. Math.*, 1950, 3: 163–170
 - [9] J. D. Cole. On a quasilinear parabolic equation occurring in aerodynamics[J]. *Quart. Appl. Math.*, 1951, 9: 201–230
 - [10] M. K. Kadalbajoo, A. Awasthi. A numerical method based on Crank-Nicolson scheme for Burgers' equation[J]. *Appl. Math. Comput.*, 2006, 182: 1430–1442
 - [11] K. Pandey, L. Verma, A. K. Verma. On a finite difference scheme for Burgers' equation[J]. *Appl. Math. Comput.*, 2009, 215: 2206–2214
 - [12] S. Kutluay, A. R. Bahadir, A. özdeş. Numerical solution of one-dimensional Burgers equation; explicit and exact-explicit finite difference methods[J]. *J. Comput. Appl. Math.*, 1999, 103: 251–261
 - [13] W. Y. Liao. An implicit fourth-order compact finite difference scheme for one-dimensional Burgers' equation[J]. *Appl. Math. Comput.*, 2008, 206: 755–764
 - [14] S. Dhawan, S. Kapoor, S. Kumar, S. Rawat. Contemporary review of techniques for the solution of nonlinear Burgers equation[J]. *J. Comput. Sci.*, 2012, 3(5): 405–419
 - [15] T. A. Elgohary, L. Dong, J. L. Junkins, S. N. Atluri. Time Domain Inverse Problem in Nonlinear Systems Using Collocation & Radial Basis Functions[J]. *CMES; Comput. Model. Eng. Sci.*, 2014, 100(1): 59–84
 - [16] B. Enquist, A. Majda. Absorbing boundary conditions for numerical simulation of waves[J]. *Math. Comput.*, 1977, 31(139): 629–651
 - [17] K. Feng. Finite element method and natural boundary reduction[J]. *Proc. of Inter. Cong. Math.*, 1983: 1439–1453
 - [18] D. Givoli. *Numerical Methods for Problem in Infinite Domains*[M]. Amsterdam: Elsevier, 1992
 - [19] H. D. Han, Z. Y. Huang. A class of artificial boundary conditions for heat equation in unbounded domains[J]. *Comput. Math. Appl.*, 2002, 43(6–7): 889–900
 - [20] H. D. Han, Z. Y. Huang. Exact and approximating boundary conditions for the parabolic problems on unbounded domains[J]. *Comput. Math. Appl.*, 2002, 44(5–6): 655–666
 - [21] H. D. Han, X. N. Wu. *Artificial Boundary Method*[M]. Beijing: Tsinghua University Press/Spring Press, 2012
 - [22] H. D. Han, X. N. Wu, Z. L. Xu. Artificial boundary method for Burgers' equation using nonlinear boundary conditions[J]. *J. Comput. Math.*, 2006, 24(3): 295–304
 - [23] B. Hosseini, R. Hashemi. Solution of Burgers equation using a Local RBF Meshless Method[J]. *Inter. J. Comput. Meth. Eng. Sci. & Mech.*, 2011, 12(1): 44–58

- [24] D. J. Liu, D. H. Yu. The coupling method of natural boundary element and mixed finite element for stationary Navier-Stokes[J]. CMES: Comput. Model. Eng. Sci. , 2008, 37(3): 305—329
- [25] Z. Z. Sun, X. N. Wu. A difference scheme for Burgers equation in an unbounded domain[J]. Appl. Math. Comput. , 2009, 209(2): 285—304
- [26] X. N. Wu, Z. Z. Sun. Convergence of difference scheme for heat equation in unbounded domains using artificial boundary conditions[J]. Appl. Numer. Math. , 2004, 50(2): 261—277
- [27] X. N. Wu, J. W. Zhang. High-order local absorbing boundary conditions for heat equation in unbounded domains [J]. J. Comput. Math. , 2011, 29(1): 74—90
- [28] D. H. Yu. Natural Boundary Integral Method and Its Applications[M]. Beijing/Dordrecht/New York/London: Kluwer Academic Publisher/Science Press, 2002
- [29] D. H. Yu, H. Y. Huang. The artificial boundary method for a nonlinear interface problem on unbounded domain [J]. CMES: Comput. Model. Eng. Sci. , 2008, 35(3): 227—252
- [30] T. Zhang, Y. He, L. Dong, S. Li, A. Alotaibi, S. N. Atluri. Meshless Local Petrov-Galerkin Mixed Collocation Method for solving Cauchy inverse problems of steady-state heat transfer[J]. CMES: Comput. Model. Eng. Sci. , 2014, 97(6): 509—533
- [31] Q. Zheng, J. Wang, J. Y. Li. The coupling method with the natural boundary reduction on an ellipse for exterior or anisotropic problems[J]. CMES: Comput. Model. Eng. Sci. , 2011, 72(2): 103—113

High-precision numerical methods for solving nonlinear Burgers' equation

Fan Lei

(School of Economics and Management, North China Univ. of Tech. , Beijing, 100144)

Abstract Burgers' equation is an important and basic parabolic PDE in fluid mechanics. Various numerical schemes for solving Burgers' equation can be summarized as finite difference, finite element, spectral element, spline approximations and so on. In this paper, the main content is divided into two parts. We study numerical methods for Burgers' equation in bounded and unbounded domain respectively.

In the first part, we study the numerical solution of one-dimensional Burgers' equation with non-homogeneous Dirichlet boundary conditions in bounded domain. This nonlinear problem is converted into the linear heat equation with non-homogeneous Robin boundary conditions by Hopf-Cole transformation. The heat equation is discretized by Crank-Nicolson finite difference scheme, and the fourth-order difference schemes for the Robin conditions are combined with the Crank-Nicolson scheme at two endpoints. The proposed method is proved to be second-order convergent and unconditionally stable. The numerical example supports the theoretical results.

In the second part, we construct a numerical method for one-dimensional Burgers' equation in the unbounded domain by using artificial boundary conditions. The original problem is converted by Hopf-Cole transformation to the heat equation in the unbounded domain, the latter is reduced to an

equivalent problem in a bounded computational domain by using two artificial integral boundary conditions, a finite difference method with discrete artificial boundary conditions is established by using the method of reduction of order for the last problem, and thereupon the numerical solution of Burgers' equation is obtained. This artificial boundary method is proved and verified to be uniquely solvable, unconditionally stable and convergent with the order 2 in space and the order $3/2$ in time for solving Burgers' equation on the computation domain.

Key words burgers' equation; hopf-cole transformation; crank-nicolson scheme; unbounded domain; artificial boundary method; finite difference method

非线性奇异摄动对流扩散方程的高精度数值解法

研究生 李雪峰

导师 郑 权

(北方工业大学理学院 北京 100144)

摘 要 本文我们主要将非线性奇异摄动对流扩散问题分为以下两部分进行研究:

(1)第一部分研究一类混合差分格式在 B-S 网络上求解非线性奇异摄动对流扩散方程,首先,构造二阶混合差分格式,在边界层部分用中心差商格式,非边界层部分为中点迎风差分格式。然后,引入离散线性算子,证明其为 M 矩阵,并结合比较原则,得到二阶收敛的结果($O(N^{-2})$)。而且,我们对差分格式的组合系数做了扩展,在最后的数值算例中进行了比较,数值算例支持了我们的理论结果。

(2)第二部分在第一部分的基础上,讨论在混合差分格式上的导数逼近,利用解的误差估计的证明方法,通过使用向后差分,获得导数在 $x_i - \frac{1}{2}$ 处的误差估计,结论为二阶收敛($O(N^{-2})$)。其次,又对已有著作中的 S 网格上的误差估计进行了改善,得到近于二阶收敛的结果($O(N^{-2} \ln^2 N)$)。数值算例对 S 网和 B-S 网的误差结果进行了比较,且证实了我们的理论结果。

关 键 词 非线性奇异摄动对流扩散; 层适应; 导数逼近; 混合格式; 一致收敛

作者简介 李雪峰,女,1989 年出生,硕士研究生,主要研究方向为微分方程数值解法。

E-mail: 1011510358@qq.com

1 绪论

奇异摄动问题的出现最早是在研究天体力学行星运动和流体力学粘性运动中,后来在半导体设备模型和金融模型等方面也都出现了此类方程,在这里最著名的是高雷诺数的纳维-斯托克斯方程。由于奇异摄动问题在科学研究和工程实践中的广泛应用,如何用数值方法求解奇异摄动问题成为越来越多物理学家和数学家关注的焦点。因此,在各相关领域科研工作者的共同努力下,奇异摄动问题的数值解法得到了快速发展,并且获得了很多好的成果,大量论文和专著也相继发表,奇异摄动问题已成为应用数学的一个重要分支,成为解决非线性数值问题的一种重要方法,并且已经成为国际上科学界研究的重要课题之一。

由于奇异摄动对流扩散问题具有的广泛物理背景和实际意义,它的数值解法的研究成为一种必然趋势。

奇异摄动方程中存在着小的摄动参数,因此在求解方程中会出现边界层和内部层,并且其真解在边界层部分的变化会很大。为了解决此类问题,数值解法的选取至关重要,经典的方法主要包括有限元法、边界元法、有限差分法、有限体积法和谱方法。为了提高求解过程中的精度,差分格式的构造同样至关重要,其中差分格式包括简单迎风格式、中点迎风格式和中心差商格式。

对于一般的线性问题,用均匀网格求解便可以得到很好的数值结果,然而对于解决奇异摄动对流扩散问题,在任意网格和均匀网格上求解都是不现实的,其缺点是计算量过大,远远超出了计算机的硬件能力。

而最主要的原因是奇异摄动对流扩散问题是存在边界层的,此类问题在边界层部分变化剧烈,非边界层部分则相对光滑。若要采用均匀网格,对网格的划分要求便会很严格,因此需要采用细密的网格划分,不然会对我们所求的误差产生较大影响,为增加误差精度,对边界层部分采用细网格,平滑部分采用粗网格,这样既能减小计算量又能保证误差精度。

N. S. Bakhvalov[1] 早在 60 年代对流扩散问题中提出了边界层适应法,并通过引入网格生成函数构造特殊网格求解奇异摄动问题。八十年代中后期, E. C. Gartland[2] 和 R. Vulanović[3] 利用特殊网格,构造了奇异摄动问题的差分格式,得到了关于 ϵ 是一致收敛的数值解。对半线性的奇异摄动问题, G. Sun 和 M. Stynes[4] 在 1995 年提出了高阶收敛性差分格式。H. —G. Ross 和 T. Linß[5] 在 1999 年得到了关于网格特征函数的一些条件,这些条件对于方法收敛是充分的,关于摄动参数是均匀的,这些条件很容易验证并且很快推断得出收敛速度。

近年来,人们致力于用非均匀网格来解决奇异摄动对流扩散问题,其中最经典的非均匀网格构造方法有两种:层适应网格和自适应网格。其中层适应网格包括 Bakhvalov 网格、Shishkin 网格和人们根据 Bakhvalov 网格与 Shishkin 网格的利弊构造出的 Bakhvalov-Shishkin 网格等。

在过去一些年内,解决线性奇异摄动对流扩散问题有大量的数值方法,包括:在层适应网格下的简单迎风格式、中点迎风格式[7–12]。其中, M. Stynes 和 H. —G. Roos[11] 在 Shishkin 网络上运用中点迎风差分格式分析了奇异摄动两点边值问题,在细网格部分利用中心差商格式和在粗网格部分利用中点迎风格的混合格式,证明了该格式优于标准的迎风差分格式,获得了二阶收敛。而且, N. V. Kopteva[12] 对差分格式选择了自由的组合权重。

随着科学的进步与发展,我们遇到的问题除了线性问题,还存在着大量的非线性问题非线性方程在实际问题中经常出现,并且在工程与科学计算中我们熟悉的很多线性模型都是通过非线性问题转化的,因此,由于非线性方程的复杂性、广泛性以及非线性奇异摄动方程在工程和物理中的应用,研究非线性奇异摄动方程很有意义。因此基于线性奇异摄动方程的研究成果,我们把它推广并应用到非线性奇异摄动对流扩散方程中去。

非线性奇异摄动对流扩散方程:

$$\begin{cases} Tu(x) := -\epsilon u''(x) - b(x, u(x))' + c(x, u) = 0, & x \in (0, 1), \\ u(0) = u(1) = 0, \end{cases}$$

其中 $0 < \epsilon < 1$ 是摄动参数,其中 ϵ 是一个很小的摄动参数并且满足 $0 < \epsilon < 1$, 当 $x \in [0, 1], u \in R$ 时, $0 < \beta \leq b_u(x, u) \leq \beta^*, 0 \leq c_u(x, u) \leq \gamma^*, \beta, \beta^*, \gamma^*$ 是常数。方程有唯一的解并且在 $x=0$ 处有指数边界层,解 u 以及其任意阶导数 q 满足 $|u^{(k)}(x)| \leq C(1 + \epsilon^{-k} e^{-\beta x/\epsilon}), x \in [0, 1]$, 其中 $k=0, 1, \dots, q$ 。解 u 的性质最先由 R. B. Kellogg 和 A. Tsan[6] 对线性问题证明成立,然后由 R. Vulanović[12] 推广到非线性奇异摄动问题上。而对于 $c(x, u(x)) = -f(x)$ 的非线性问题,用最大模的一阶、二阶差分格式的后验误差估计由 N. Kopteva[13] 在 2001 年的文章中获得。T. Linß 在[14] 研究了简单迎风格式和二阶混合差分格式。在 2000 年, T. Linß 等[15] 用一阶差分格式证明了误差是 (l_∞, l_1) 稳定的,在[16] 中,则利用二阶混合差分格式研究了最大误差。

除了研究解的误差估计之外,导数的误差估计在线性奇异摄动对流扩散问题上被[17, 18] 研究。带权导数在子区间中点的误差在 S 型网格[19] 中被得到。对于线性守恒型对流扩散问题,由 (l_∞, l_1) 稳定性,带权和不带权的导数逼近在迎风差分格式[20] 上是一阶收敛的。近年来,通过在迎风差分格式上利用网格等分布,得到了一阶收敛[21]。众所周知,对流扩散方程的高精度数值解法的获得并没有自动地得到好的导数逼近。而且,导数逼近在很多领域也有所应用[22, 21, 19, 23]。

2 非线性奇异摄动对流扩散方程的混合差分格式

我们考虑如下非线性奇异摄动对流扩散方程的边值问题:

$$\begin{cases} Tu(x) := -\epsilon u''(x) - b(x, u(x))' + c(x, u) = 0, & x \in (0, 1), \\ u(0) = u_0, u(1) = u_1, \end{cases} \quad (2-1)$$

其中 ϵ 是小的摄动参数并满足 $0 < \epsilon \ll 1$, b 和 c 是充分光滑的函数, 当 $x \in (0, 1)$, $u \in R$ 时, $0 < b_u(x, u) \leq \beta^*$, $0 \leq c_u(x, u) \leq \gamma^*$, 并且 $\beta, \beta^*, \gamma^*, u_0, u_1$ 都是常数。此方程有唯一解 u , 且指数边界层在 $x=0$ 处。解 u 和它的任意阶导数 q 满足

$$|u^{(k)}(x)| \leq C(1 + \epsilon^{-k} e^{-\frac{\beta x}{\epsilon}}), x \in (0, 1) \quad (2-2)$$

其中 $k=0, 1, \dots, q$ (见[24])。

2.1 网格函数的构造

为解决指数边界层, 我们利用如下 B-S(Bakhvalov-Shishkin) 网格: $x_i = x(t_i)$, $t_i = \frac{i}{N}$, $i=0, 1, \dots, N$, 网格函数是:

$$x(t) = \begin{cases} -\frac{\alpha\epsilon}{\beta} \ln(1 - 2(1 - N^{-1})t), & 0 \leq t \leq \frac{1}{2}, \\ 1 - 2(1 - \frac{\alpha\epsilon}{\beta} \ln N)(1 - t), & \frac{1}{2} \leq t \leq 1. \end{cases} \quad (2-3)$$

其中离散参数 N 是正整数, $\alpha \geq 2$, 转折点 $x_{\frac{N}{2}} = \frac{\alpha\epsilon}{\beta} \ln N$, ϵ 满足 $\epsilon \leq CN^{-1}$, 在此文章中, C 是与 ϵ 和 h_i ($h_i = x_i - x_{i-1}$) 无关的正常数, 并记 C 在不同情况下有不同的值。

2.2 Bakhvalov-Shishkin 网格的性质

引理 2.1 h_i 和 $h_i + \frac{N}{2}$ 单调递增, 且满足: $\frac{2\alpha\epsilon(N-1)}{\beta N^2} < h_i < \frac{2\alpha\epsilon}{\beta}$ 且 $N^{-1} \leq h_i + \frac{N}{2} = H \leq 2N^{-1}$, $i=1, 2, \dots, \frac{N}{2}$, 其中 $H = \frac{1 - \alpha\epsilon \frac{\ln N}{\beta}}{\frac{N}{2}}$ 。并且, $h_{\frac{N}{2}} - i < \frac{2\alpha\epsilon}{(2i+1)\beta} < \frac{2\alpha\epsilon}{3\beta}$, $i=1, 2, \dots, \frac{N}{2}$, 当 $N \rightarrow \infty$ 时, $h_{\frac{N}{2}} \rightarrow \frac{2\epsilon}{\beta}$ 。

2.3 混合差分格式

对于方程(1-1), 我们考虑如下混合差分格式:

$$[T^N u^N]_i = \frac{[A^N u^N]_{i+1} - [A^N u^N]_i}{h_{\sigma, i}} = 0, i=1, 2, \dots, N-1, \quad (2-4)$$

其中

$$[A^N u^N]_i := \epsilon \frac{u_i - u_{i-1}}{h_i} + \sigma_i b(x_i, u_i) + (1 - \sigma_i) b(x_{i-1}, u_{i-1}) - \sum_{j=1}^{i-1} \sigma_j c(x_{\sigma, j} + \frac{1}{2}, u_{\sigma, j} + \frac{1}{2}),$$

$$x_{\sigma, i} = (1 - \sigma_i) x_{i-1} + \sigma_i x_i, x_{\sigma, i} - \frac{1}{2} = \frac{x_{\sigma, i} - 1 + x_{\sigma, i}}{2}, h_{\sigma, i} = (1 - \sigma_i) h_i + \sigma_{i+1} h_{i+1}, u_{\sigma, j - \frac{1}{2}} = u(x_{\sigma, j - \frac{1}{2}})$$

其中递增变量

$$\frac{1}{2} = \sigma_1 = \dots = \sigma_{\frac{N}{2} - j_1} \leq \dots \leq \sigma_{\frac{N}{2} - j_0 + 1} = \dots = \sigma_{\frac{N}{2}} = \dots = \sigma_N = 1 (0 < j_0 < j_1 < \dots < N/2), \text{ 并且当 } \sigma_i = 1$$

时, $h_i b_{i-1} > 2\epsilon$, 这里 $b_i = \int_0^1 b_u(x_i, w_i + s(v_i - w_i)) ds$ 。

对任意网格函数 $\{v_i\}$, 我们记离散算子: $\|v\| = \max_{i=0, \dots, N} |v_i|$ 。离散线性算子定义如下:

$$L^N y_i = -\frac{M^N y_i + 1 - M^N y_i}{h_{\sigma,i}}, \quad (2-5)$$

其中

$$M^N y_i = \epsilon \frac{y_i - y_{i-1}}{h_i} + \sigma_i b_i y_i + (1 - \sigma_i) b_{i-1} y_{i-1} - \sum_{j=1}^{i-1} h_{\sigma,j} \bar{c}_j y_{\sigma,j+1/2}$$

$$b_i = \int_0^1 b_u(x_i, w_i + s(v_i - w_i)) ds, \bar{c}_i = \int_0^1 c_u(x_{\sigma,i+1/2}, w_{\sigma,i+1/2} + s(v_{\sigma,i+1/2} - w_{\sigma,i+1/2})) ds.$$

明显的, 我们有 $\beta^* \geq b_i \geq \beta > 0$ 和 $\gamma^* \geq \bar{c}_i \geq 0$ 。假设 $b_i = \beta, i = 1, 2, \dots, \frac{N}{2}$, 那么有 $b_{i-1} h_i \leq \beta h_{N/2}$

$$\rightarrow 2\epsilon, i = 1, 2, \dots, \frac{N}{2} \text{ 和 } b_{i-1} h_i \leq \beta h_{N/2-1} \rightarrow \frac{4}{3}\epsilon, i = 1, 2, \dots, \frac{N}{2} - 1。$$

引理 2.2 对 $H < \frac{\beta}{4\gamma^* e^{\gamma^*/\beta}}$, 公式(1-5)的离散线性算子 M^N 的系数矩阵是 M -矩阵。

引理 2.3 对任意网格函数 $\{w_i\}$ 和 $\{v_i\}$, 有

$$\|v - w\| \leq C_{i=1,2,\dots,N-1}^{\max} \left| \sum_{j=i}^{N-1} \sigma_{\sigma,j} [T^N v - T^N w]_j \right|。$$

2.4 误差估计

引理 2.4 $\|u - u^N\|_{\infty} \leq CN^{-2} + C_{i=1,2,\dots,N-1}^{\max} \int_{x_{i-1}}^{x_{i+1}} (1 + \epsilon^{-2} e^{-\beta x/\epsilon})(x - x_{i-1}) dx。$

定理 2.1 如果引理 2.4 的条件成立, 则方程(1-1)的误差估计为:

$$\|u - u^N\|_{\infty} \leq CN^{-2}$$

证明: 对任意的 $k \in N^+$, 任意单调递减函数 $g(x)$ 在区间 $[a, b]$ 上满足不等式

$$\int_a^b g(x)(x-a)^{k-1} dx \leq \frac{1}{k} \left\{ \int_a^b g(x)^{1/k} dx \right\}^k,$$

则有

$$\int_{x_{i-1}}^{x_{i+1}} (1 + \epsilon^{-2} e^{-\beta x/\epsilon})(x - x_{i-1}) dx \leq \frac{1}{2} \int_{x_{i-1}}^{x_{i+1}} (1 + \epsilon^{-1} e^{-\beta x/2\epsilon}) dx)^2。$$

对 $i = 1, \dots, \frac{N}{2}$, 我们有

$$\begin{aligned} \int_{x_{i-1}}^{x_i} (1 + \epsilon^{-1} e^{-\beta x/2\epsilon}) dx &= h_i + \int_{x_{i-1}}^{x_i} \epsilon^{-1} e^{-\beta x/2\epsilon} dx \\ &= h_i + \frac{2}{\beta} \int_{t_{i-1}}^{t_i} (1 - 2(1 - N^{-1})t) \frac{2(1 - N^{-1})}{1 - 2(1 - N^{-1})t} dt \leq CN^{-1} \end{aligned}$$

对 $i = \frac{N}{2} + 1, \dots, N$, 我们有 $\int_{x_{i-1}}^{x_i} (1 + \epsilon^{-1} e^{-\beta x/2\epsilon}) dx = h_i - \frac{2}{\beta} e^{-\beta x/\epsilon^2} \Big|_{x_{i-1}}^{x_i} \leq CN^{-1}。$

因此我们得到结论 $\|u - u^N\|_{\infty} \leq \frac{C}{2} C_{i=1, \dots, N-1}^{\max} \left(\int_{x_{i-1}}^{x_{i+1}} (1 + \epsilon^{-1} e^{-\beta x/2\epsilon}) dx \right)^2 \leq CN^{-2}。$

3 基于混合差分格式的导数逼近

关于求解对流扩散问题的解的误差估计有大量的作者利用有限差分方法进行研究, 然而对于导数逼近却很少有作者研究, 尽管导数逼近在实际应用中被需要, 例如: 标准导数需要被用来计算摩擦系数和应力强度因子。但只有少数的文章考虑了对流扩散方程的导数逼近[18, 20, 39]。V. Ervin [18]仅

仅讨论了在边界层外的不带权的导数逼近。对于守恒型对流扩散方程, N. Kopteva [20] 用非均匀网格(分段均匀 S 网格和 B 网格)获得了在整个区间的导数逼近。而且近年来主要是对于线性对流扩散方程的迎风差分格式进行研究, 获得一阶一致收敛的结果[20-21]。因此, 本章我们主要讨论基于混合差分格式的导数逼近。

3.1 混合差分策略及导数逼近

对于方程(1-1), 我们仍考虑如下混合差分格式:

$$[T^N u^N]_i = \frac{[A^N u^N]_{i+1} - [A^N u^N]_i}{h_{\sigma,i}} = 0, i=1, 2, \dots, N-1, \quad (3-1)$$

其中

$$[A^N u^N]_i := \epsilon \frac{u_i - u_{i-1}}{h_i} + \sigma_i b(x_i, u_i) + (1 - \sigma_i) b(x_{i-1}, u_{i-1}) - \sum_{j=1}^{i-1} h_{\sigma,j} c(x_{\sigma,j} + \frac{1}{2}, u_{\sigma,j+1/2}),$$

$$x_{\sigma,i} = (1 - \sigma_i) x_{i-1} + \sigma_i x_i, \quad x_{\sigma,i-1/2} = \frac{x_{\sigma,i-1} + x_{\sigma,i}}{2}, \quad h_{\sigma,i} = (1 - \sigma_i) h_i + \sigma_{i+1} h_{i+1}, \quad u_{\sigma,j-1/2} = u(x_{\sigma,j-1/2})$$

组合系数 $\sigma_i = \begin{cases} \frac{1}{2}, & \frac{1}{2} \leq \epsilon / (h_i \beta), \\ 1, & \text{其他}, \end{cases} \quad (3-2)$

即 $\frac{1}{2} = \sigma_1 = \dots = \sigma_{\bar{j}} < \sigma_{\bar{j}+1} = \dots = \sigma_N = 1$, 其中 $\beta = \max_{0 \leq x \leq \frac{1}{2}} b_u(x, u)$, $\bar{j} = \max \left\{ i \mid \frac{1}{2} \leq \epsilon / (h_i \beta) \right\}$ 。

由此给出在 $x_{i-1/2}$ 处的数值导数, 并记作: $D^- u_i = \frac{u_i - u_{i-1}}{h_i}, i=1, \dots, N$ 。 $(3-3)$

而且, 我们可以用 $\beta_\lambda = \max_{0 \leq x \leq \lambda} b_u(x, u)$, $j_\lambda = \max \left\{ i \mid \frac{1}{2} \leq \epsilon / h_i \beta_\lambda \right\}$ 分别代替(2-5)中的 β 和 \bar{j} 使其变为另一种混合差分格式。若用 β^* 和 $j_* = \max \left\{ i \mid \frac{1}{2} \leq \epsilon / h_i \beta^* \right\}$ 分别代替 β 和 \bar{j} 则是[12]中的差分格式。

对任意网格函数 $\{v_i\}$, 我们记离散算子:

$$\|v\|_\infty = \max_{i=0, \dots, N} |v_i|, \quad \|v\|_* = \max_{i=1, 2, \dots, N-1} \left| \sum_{j=1}^{i-1} h_{\sigma,j} v_j \right|.$$

记 $\|v\|_* \leq \|v\|_\infty$, $\|D^- v\|_* \leq 2 \|v\|_\infty$ 。

令 v 和 w 是任意两个网格函数。离散线性算子定义如下:

$$L^N y_i = - \frac{M^N y_{i+1} - M^N y_i}{h_{\sigma,i}}, \quad (3-4)$$

其中

$$M^N y_i = \epsilon \frac{y_i - y_{i-1}}{h_i} + \sigma_i \bar{b}_i y_i + (1 - \sigma_i) \bar{b}_{i-1} y_{i-1} - \sum_{j=1}^{i-1} h_{\sigma,j} \bar{c}_j y_{\sigma,j+1/2}$$

$$\bar{b}_i = \int_0^1 b_u(x_i, w_i + s(v_i - w_i)) ds, \quad \bar{c}_i = \int_0^1 c_u(x_{\sigma,i+1/2}, w_{\sigma,i+1/2} + s(v_{\sigma,i+1/2} - w_{\sigma,i+1/2})) ds.$$

算子 L^N 和 T^N 有满足如下等式:

$$L^N(v - w) = T^N v - T^N w.$$

引理 3.1 对 $H < \frac{\beta}{4\gamma^* e^{\gamma^*/\beta}}$, 公式(2-4)的离散线性算子 M^N 的系数矩阵是 M -矩阵。

3.2 误差估计

引理 3.2 令 $\{y_i\}$ 为任意网格函数, 且满足 $L^N y_i = f_i, i=1, \dots, N-1, y_0 = y_N = 0$, 则 $\|y\|_\infty \leq C \|f\|_*$, $\max_{i=1, 2, \dots, N-1} \epsilon |D^- y_i| \leq C \|f\|_*$ 。

非线性奇异摄动对流扩散方程(2-1)数值解和导数数值解的一致收敛性结合(2-3), (3-3)通过如下能够被证明是 (l_∞, l_1) 稳定的。误差 $e_i^N = u_i - u_i^N$ 满足:

$$L^N e_i^N = L^N(u_i - u_i^N) = T^N u_i - T^N u_i^N = T^N u_i. \quad (3-5)$$

定理 3.1 非线性方程(2-1)的导数误差估计满足:

$$\max_{i=1, \dots, N} \epsilon |D^- u_i^N - u'_{i-1/2}| \leq CN^{-2}$$

证明:由泰勒展开式, 导数数值解误差在 $x_{i-1/2}$ 处满足

$$\begin{aligned} \max_{i=1, \dots, N} \epsilon |D^- u_i^N - u'_{i-1/2}| &\leq \epsilon (|D^- u_i^N - D^- u_i| + |D^- u_i - u'_{i-1/2}|) \\ &\leq \epsilon |D^- e_i^N| + \left| \frac{3\epsilon}{2} \int_{x_{i-1}}^{x_i} u'''(x)(x-x_{i-1}) dx \right| \end{aligned}$$

由引理 3.2, (2-2)和定理 2.1 的证明, 我们有

$$\begin{aligned} \max_{i=1, \dots, N} \epsilon |D^- u_i^N - u'_{i-1/2}| &\leq C_0 \|L^N e_i^N\|_* + \max_{i=1, \dots, N-1} \left| \int_{x_{i-1}}^{x_i} (1 + \epsilon^{-2} e^{-\beta x/\epsilon})(x-x_{i-1}) dx \right| \\ &= C_0 \|T^N u_i\|_* + \max_{i=1, \dots, N-1} \left| \int_{x_{i-1}}^{x_i} (1 + \epsilon^{-2} e^{-\beta x/\epsilon})(x-x_{i-1}) dx \right| \leq CN^{-2} \end{aligned}$$

对于标准的 Shishkin 网格, 网格函数为:

$$x(t) = \begin{cases} \frac{2\alpha\epsilon \ln N}{\beta} t, & 0 \leq t \leq \frac{1}{2} \\ 1 - 2(1 - \frac{\alpha\epsilon}{\beta} \ln N)(1-t), & \frac{1}{2} \leq t \leq 1 \end{cases} \quad (3-6)$$

引理 3.4 $h_i = \frac{2\alpha\epsilon \ln N}{\beta N}$, $N^{-1} \leq h_{N/2+i} \leq 2N^{-1}$, $i=1, 2, \dots, \frac{N}{2}$ 。

当 $\frac{N}{\ln N} \geq \alpha \frac{\beta}{\beta}$ 时, 由(3-2), 有组合系数 $\frac{1}{2} = \sigma_1 = \dots = \sigma_{N/2} < \sigma_{N/2+1} = \dots = \sigma_{N-1} = 1$, 其中 β 可以用 β^*

和 β_λ 代替。因此, 我们有:

定理 3.2 当 $\frac{N}{\ln N} \geq \alpha \frac{\beta}{\beta}$ 时(或 $\alpha \frac{\beta^*}{\beta}$ 和 $\alpha \frac{\beta_\lambda}{\beta}$), (2-4) - (2-6)在 S 网格上解和导数的误差估计分

别满足:

$$\begin{aligned} \|u - u^N\|_\infty &\leq CN^{-2} \ln^2 N \\ \max_{i=1, \dots, N} \epsilon |D^- u_i^N - u'_{i-1/2}| &\leq CN^{-2} \ln^2 N. \end{aligned}$$

证明:由定理 2.1 的证明和(3-6), 对 $j > \frac{N}{2}$, 因为 $\sigma_j = 1$, 由 u 在 x_j 处的泰勒展开式, 我们有:

$$[B^N u]_j - (Bu)(x_{\sigma, j}) = \epsilon \left(\frac{u_j - u_{j-1}}{h_j} - u'_j \right) = \frac{\epsilon}{h_j} \int_{x_{j-1}}^{x_j} u''(x)(x-x_{j-1}) dx. \text{ 所以, (2-10)成立. 对 } j \leq \frac{N}{2},$$

有 $\sigma_j = \frac{1}{2}$, (2-10)依然成立。类似于定理 2.1 的证明, 我们得到

$$\|u - u^N\|_\infty \leq C_{i=1, \dots, N-1} \left(\int_{x_{i-1}}^{x_{i+1}} (1 + \epsilon^{-1} e^{-\beta x/2\epsilon}) dx \right)^2.$$

对 $i=1, \dots, \frac{N}{2}$, 我们有

$$\int_{x_{i-1}}^{x_i} (1 + \epsilon^{-1} e^{-\beta x/2\epsilon}) dx = h_i + \frac{2\alpha \ln N}{\beta} \int_{t_{i-1}}^{t_i} e^{\alpha t \ln N} dt = h_i \frac{2}{\beta} e^{-\alpha t \ln N} \Big|_{t_{i-1}}^{t_i} \leq CN^{-1} \ln N.$$

对 $i = \frac{N}{2} + 1, \dots, N$, 我们有 $\int_{x_{i-1}}^{x_i} (1 + \epsilon^{-1} e^{-\beta x/2\epsilon}) dx = h_i - \frac{2}{\beta} e^{-\beta x/2\epsilon} \Big|_{x_{i-1}}^{x_i} \leq CN^{-1}.$

因此,我们得到 $\|u - u^N\|_\infty \leq CN^{-2} \ln^2 N$ 。

同样的,根据定理 2.2 的证明方法,我们得到在 S 网格上的导数误差估计为:

$$\max_{i=1,\dots,N} \epsilon |D^- u_i^N - u'_{i-1/2}| \leq CN^{-2} \ln^2 N。$$

3.3 数值实验

为了证实方程 (2-1) 的误差是二阶收敛的,在数值算例中我们用公式: $rate = \log_2 \frac{\max |u_i - u_i^N|}{\max |u_i - u_i^N| N_i}$ 来计算收敛率。用公式 $const = \frac{\max |u_i - u_i^N|}{N^{-2}}$ 来计算在 $B-S$ 网格上的误差估计常数。用公式 $const = \frac{\max |u_i - u_i^N|}{N^{-2} \ln^2 N}$ 来计算在 S 网格上的误差估计常数。

例 3.1 考虑线性对流扩散问题(见[20]):

$$\begin{cases} -\epsilon u'' - [(1+x)^3 u]' + \left[3(1+x)^2 + \frac{12\epsilon}{(1+x)^2} \right] u + \left[\frac{\epsilon}{4} - 3(x+1)^2 - \frac{12\epsilon}{(x+1)^2} \right] e^{\frac{x}{2}} = 0, x \in (0, 1), \\ u(0) = 2, u(1) = \frac{1}{8} e^{-\frac{15}{4\epsilon}} + e^{-\frac{1}{2}}. \end{cases}$$

准确解为 $u = \frac{1}{(1+x)^3} e^{\frac{1}{4\epsilon} - \frac{(x+1)^4}{4\epsilon}} + e^{-\frac{x}{2}}$ 。

因为 $b_u = (1+x)^3$, 我们有 $\beta = 1$, 其他三种混合格式分别为 $\beta^* = 8, \beta = \frac{27}{8}, \beta_\lambda = (1+\lambda)^3$ 。

表 3-1 比较了在 $B-S$ 网和 S 网上的误差,其中混合格式为 β^* , 同样,表 3-2 比较的是导数误差。表 3-3 和 3-4 分别是误差和导数误差,其中有 β^*, β_λ 和平滑的格式四种比较对比。从每一个表格中,我们都可看出,在 $B-S$ 网上为二阶收敛($O(N^{-2})$), S 网近于二阶收敛($O(N^{-2} \ln^2 N)$)。

表 3-1 解的误差, $\beta^* = 8$

Method	N	$\epsilon = 10^{-4}$				$\epsilon = 10^{-6}$			
		j^*	$\ u^N - u\ _\infty$	rate	const	j^*	$\ u^N - u\ _\infty$	rate	const
B-S mesh	32	9	1.115209e-02	1.90	11.4197	9	1.116705e-02	1.90	11.4351
	64	24	3.335438e-03	1.74	13.6620	24	3.342886e-03	1.74	13.6925
	128	56	8.518338e-04	1.97	13.9564	56	8.551868e-04	1.97	14.0114
	256	120	2.150407e-04	1.99	14.0929	120	2.163929e-04	1.98	14.1815
	512	248	5.392440e-05	2.00	14.1360	249	5.443151e-05	1.99	14.2689
S mesh	32	1	6.663793e-02	0.53	0.6153	1	6.665305e-02	0.53	0.6154
	64	1	4.321401e-02	0.62	0.6650	1	4.322100e-02	0.62	0.6651
	128	64	7.190769e-04	5.90	0.5004	64	7.190418e-04	5.91	0.5004
	256	128	2.334662e-04	1.62	0.4976	128	2.334310e-04	1.62	0.4975
	512	256	7.363673e-05	1.66	0.4960	256	7.361513e-05	1.66	0.4959

表 3-2 在 $x_{i-1/2}$ 处的导数误差

Method	N	$\varepsilon = 10^{-4}$			$\varepsilon = 10^{-6}$		
		$\max_{1 \leq i \leq N} \left\ D^- u_i^N - u'_{i-\frac{1}{2}} \right\ $	rate	const	$\max_{1 \leq i \leq N} \left\ D^- u_i^N - u'_{i-\frac{1}{2}} \right\ $	rate	const
B-S	32	1.907521e-02	1.82	19.5330	1.907228e-02	1.82	19.5300
	64	5.877725e-03	1.70	24.0752	5.879132e-03	1.70	24.0809
	128	1.495176e-03	1.97	24.4970	1.497130e-03	1.97	24.5290
	256	3.762753e-04	1.99	24.6596	3.773995e-04	1.99	24.7333
	512	9.421573e-05	2.00	24.6981	9.472101e-05	1.99	24.8305
S	32	1.070933e-01	0.09	0.9888	1.070964e-01	0.09	0.9888
	64	8.434551e-02	0.34	1.2980	8.434672e-02	0.34	1.2980
	128	2.559746e-03	5.04	1.7814	2.559616e-03	5.04	1.7813
	256	8.787942e-04	1.54	1.8730	8.787257e-04	1.54	1.8728
	512	2.864336e-04	1.62	1.9294	2.863975e-04	1.62	1.9292

表 3-3 解的误差 ($\varepsilon = 10^{-6}$)

smooth			β^*			$\bar{\beta}$			β_λ			
method	N	$\ u^N - u\ _\infty$	rate	j_*	$\ u^N - u\ _\infty$	rate	\bar{j}	$\ u^N - u\ _\infty$	rate	j_λ	$\ u^N - u\ _\infty$	rate
E-S												
mesh												
$\alpha = 2$	32	4.7294e-3	1.85	9	1.1167e-2	1.90	13	4.0312e-3	1.80	15	1.7331e-3	2.05
	64	1.4716e-3	1.68	24	3.3429e-3	1.74	29	1.0830e-3	1.90	31	4.2056e-4	2.04
	128	3.7968e-4	1.95	56	8.5519e-4	1.97	61	2.8081e-4	1.95	63	1.0327e-4	2.03
	256	9.6458e-5	1.98	120	2.1639e-4	1.98	125	7.1498e-5	1.97	127	2.5565e-5	2.01
	512	2.4310e-5	1.99	248	5.4431e-5	1.99	253	1.8038e-5	1.99	255	6.3588e-6	2.01
S mesh												
$\alpha = 2$	32	4.6607e-3	1.76	1	6.6653e-2	0.53	16	6.0648e-3	3.99	16	6.0648e-3	1.41
	64	2.2473e-3	1.05	1	4.3221e-2	0.62	32	2.1435e-3	1.50	32	2.1435e-3	1.50
	128	7.1904e-4	1.64	64	7.1904e-4	3.91	64	7.1904e-4	1.58	64	7.1904e-4	1.58
	256	2.3343e-4	1.62	128	2.3343e-4	1.62	128	2.3343e-4	1.62	128	2.3343e-4	1.62
	512	7.3615e-5	1.66	256	7.3615e-5	1.66	256	7.3615e-5	1.66	256	7.3615e-5	1.66

表 3-4 导数误差 ($\varepsilon=10^{-6}$)

smooth		β^*		$\bar{\beta}$		β_λ	
method	N	$\varepsilon\ D^-u^N-u'\ _\infty$	rate	$\varepsilon\ D^-u^N-u'\ _\infty$	rate	$\varepsilon\ D^-u^N-u'\ _\infty$	rate
E-S							
$\alpha=2$	32	3.5891e-3	1.60	1.9072e-2	1.82	5.9635e-3	1.87
	64	1.1068e-3	1.70	5.8791e-3	1.70	1.5429e-3	1.95
	128	2.8362e-4	1.96	1.4971e-3	1.97	3.9162e-4	1.98
	256	7.1594e-5	1.99	3.7740e-4	1.99	9.8610e-5	1.99
	512	1.7973e-5	1.99	9.4721e-5	1.99	2.4738e-5	2.00
S							
$\alpha=2$	32	1.0710e-1	0.09	1.0710e-1	0.09	1.7052e-2	2.74
	64	8.4347e-2	0.34	8.4347e-2	0.34	6.9371e-3	1.30
	128	2.5596e-3	5.04	2.5596e-3	5.04	2.5596e-3	1.44
	256	8.7873e-4	1.54	8.7873e-4	1.54	8.7873e-4	1.54
	512	2.8640e-4	1.62	2.8640e-4	1.62	2.8640e-4	1.62

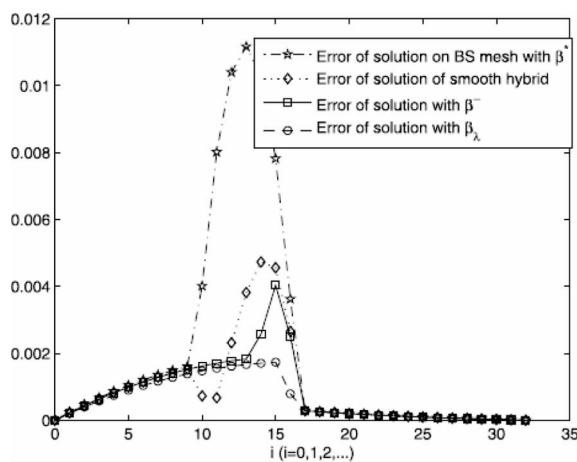


图 3-1

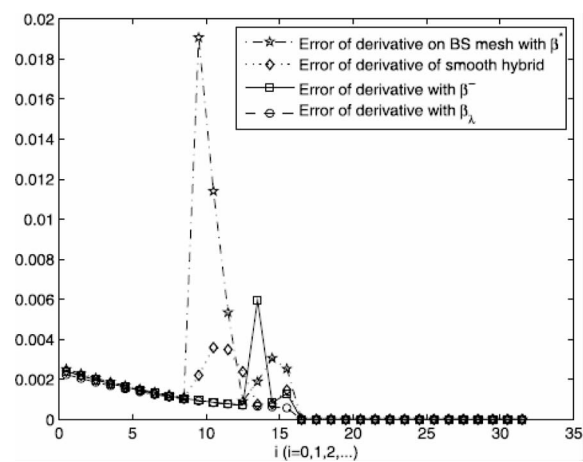


图 3-2

图 3-1 和图 3-2 分别表示在 B-S 网格上的解的误差和导数的误差,其中 $\varepsilon=10^{-6}$ 。

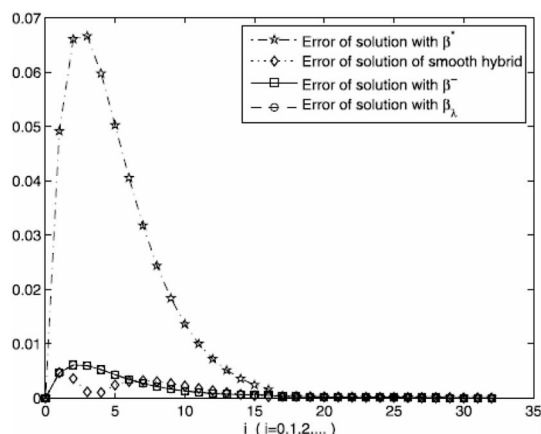


图 3-3

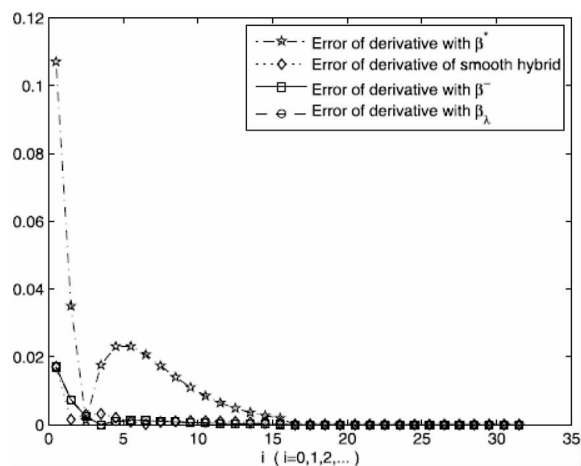


图 3-4

图 3-3 和图 3-4 分别表示在 S 网格上的解的误差和导数的误差,其中 $\epsilon=10^{-6}$ 。

数值实验结果清晰地显示,解的误差和导数误差在 B-S 网格上是二阶收敛($O(N^{-2})$),并且从两个数值算例的表格和图形中可以看出,选择更好的组合系数,会达到更小的误差。如果 N 足够大,那么混合差分格式在 S 网格上,将会接近达到二阶收敛($O(N^{-2} \ln^2 N)$)。从每一个表格的收敛率可以看出,B-S 网格的收敛率很好的达到二次收敛,而 S 网格的收敛率是介于 1 和 2 之间。因此,我们的数值算例支持理论结果。

参 考 文 献

- [1] Bakhvalov, A. S. On the optimization of methods for solving boundary value problem with boundary layers (in Russian). Zh. Vychisl. Mat. Mat. Fiz., 1969,9(4):841—859
- [2] E. C. Gartland, Graded-mesh difference schemes for singularly perturbed to point boundary value problems, Math. Comp., 1988, (51):631—657
- [3] R. Vulcanovi?, Non-equidistant generalizations of the Gushchin-Shchennikov scheme, Z. Math. Mec., 1987, (67): 625—632
- [4] G. Sunand M. Stynes, An almost fourth order uniformly convergent difference scheme for a semi-linear singularly perturbed reaction-diffusion problem. Numer. Math., 1995, (70):487—500
- [5] H. —G. Roos, T. Linß. Sufficient conditions for uniform convergence on layer adapted grids, Computing., 1999, 63 (1):27 — 45
- [6] R. B. Kellogg, A. Tsan, Analysis of some difference approximations for a singular perturbation problem without turning points, Math. Comput., 1978, (32): 1025—1039
- [7] A. S. Bakhvalov, On the optimization of methods for solving boundary value problems in the presence of boundary layers, Zh. Vychisl. Mat. Mat. Fiz. 1969, (9):841—859 (in Russian)
- [8] G. I. Shishkin, Grid approximation of singularly perturbed elliptic and parabolic equations, Second doctoral thesis, Keldysh Institute, Moscow, 1990 (in Russian)
- [9] T. Linß, An upwind difference scheme on a novel Shishkin-type mesh for a linear convection-diffusion problem, J. Comput. Appl. Math., 1999, (110):93—104
- [10] M. Stynes, H. —G. Roos, The midpoint upwind scheme, Appl. Numer. Math. 1997, (23) 361 — 374

-
- [11] V.B. Andreev, N. V. Kopteva, On the convergence, uniform with respect to a small parameter, of monotone three-point difference schemes, Differ. Uravn., 1998, (34): 921—928 (in Russian); translation in Differential Equations, 1998, (34) 921—929 (in English)
 - [12] R. Vulanović, A uniform numerical method for quasilinear singular perturbation problems without turning points, Computing, 1989, (41): 97 - 106
 - [13] N. Kopteva, Maximum norm a posteriori error estimates for a one-dimensional convection-diffusion problem, SIAM J. Numer. Anal., 2001, (39): 423—441
 - [14] T. Linß, Sufficient conditions for uniform convergence on layer-adapted grids, Appl. Numer. Math. 2001, (37): 241 - 255
 - [15] T. Linß, H. —G. Ross, R. Vulanović, Uniform pointwise convergence on Shishkin-type meshes for quasi-linear convection-diffusion problem, SIAM J. Numer. Anal. 2000, (38): 897 - 912
 - [16] T. Linß, Uniform second order pointwise convergence of a finite difference discretization for a quasilinear problem, Comput. Math. Math. Phys. 2001, (41): 898 - 909
 - [17] V.B. Andreev, The Green function and a priori estimates for solutions of monotone three-point singularly perturbed difference schemes, Differ. Equ. 2001, (37): 923 - 933
 - [18] V. Ervin, W. Layton, On the approximation of derivatives of singularly perturbed boundary value problems, SIAM J. Sci. Stat. Comput. 1987, (8): 265 - 277
 - [19] T. Linß, H. —G. Ross, Gradient recovery for singularly perturbed boundary value problems I: one-dimensional convection - diffusion, Computing 2001, (66): 163 - 178
 - [20] N. Kopteva, M. Stynes, Approximation of derivatives in a convection - diffusion two-point boundary value problem, Appl. Numer. Math. 2001, (39): 47 - 60
 - [21] J. Mohapatra, S. Natesan, Parameter-uniform numerical method for global solution and global normalized flux of singularly perturbed boundary value problems using grid equidistribution, Comput. Math. Appl. 2010, (60): 1924 - 1939
 - [22] G. I. Shishkin, Discrete Approximation of Singularly Perturbed Elliptic and Parabolic Equations, Russian Academy of Sciences, Ural Section, Ekaterinburg, 1992 (in Russian)
 - [23] R. Mythili Priyadharshini, N. Ramanujam, V. Shanthi, Approximation of derivative in a system of singularly perturbed convection - diffusion equations, J. Appl. Math. Comput. 2009, (29): 137 - 151
 - [24] R. Vulanović, A uniform numerical method for quasilinear singular perturbation problems without turning points, Computing. 1989, (41): 97 - 106
 - [25] R. E. O'Malley Jr., Introduction to Singular Perturbations. New York, London: Academic Press, 1974
 - [26] J. Lorenz, Stability and monotonicity properties of stiff quasi-linear boundary problems. Univ. u Novom Sadu Zb. rad. Prirod.-Mat. Fak. Ser. Mat., 1982, (12): 151—175
 - [27] N. S. Bakhvalov, On the optimization of methods for solving boundary value problems with boundary layers, Ž. Vyčisl. Mat. Mat. Fiz. 1969 (9): 841 - 859 (in Russian)
 - [28] G. I. Shishkin, Discrete Approximation of Singularly Perturbed Elliptic and Parabolic Equations, Russian Academy of Sciences, Ural Section, Ekaterinburg, 1992 (in Russian)
 - [29] J. J. H. Miller, R. E. O'Riordan, G. I. Shishkin, Fitted Numerical Methods for Singular Perturbation Problems. World Scientific, Singapore, 1996
 - [30] H. —G. Roos, M. Stynes, L. Tobiska, Robust Numerical Methods for Singularly Perturbed Differential Equations, Springer-Verlag, Berlin, Heidelberg, 2008
 - [31] M. K. Kadalbajoo, V. Gupta, A brief survey on numerical methods for solving singularly perturbed problems, Appl. Math. Comput. 2010, (217): 3641—3716

- [32] V.B. Andreev, N. V. Kopteva, On the convergence, uniform with respect to a small parameter, of monotone three-point difference schemes, *Differ. Uravn.*, 1998, (34): 921–928 (in Russian); translation in *Differential Equations*, 1998, (34): 921–929 (in English)
- [33] C. de Boor, Good approximation by splines with variable knots, in: A. Meir, A. Sharma (Eds.), *Spline Functions and Approximation Theory*, Proceedings of the Symposium Held at the University of Alberta, Edmonton, May 29 – June 1, 1972, Birkhäuser, Basel, 1973
- [34] A. Fröhner, T. Linß, H. — G. Roos, Defect correction on Shishkin-type meshes, *Numer. Algorithms* 2001, (26): 281 – 299
- [35] N. Kopteva, Maximum norm a posteriori error estimates for a one-dimensional convection – diffusion problem, *SIAM J. Numer. Anal.* 2001, (39): 423 – 441
- [36] N. V. Kopteva, T. Linß, Uniform second-order pointwise convergence of a central difference approximation for a quasilinear convection – diffusion problem, *J. Comput. Appl. Math.* 2001 (137): 257 – 267
- [37] T. Linß, *Layer-Adapted Meshes for Reaction – Convection – Diffusion Problems*, Lecture Notes in Mathematics, vol. 1985, Springer-Verlag, Berlin, 2010
- [38] H. — G. Roos, M. Stynes, L. Tobiska, *Robust Numerical Methods for Singularly Perturbed Differential Equations*, Springer-Verlag, Berlin, Heidelberg, 2008
- [39] P. A. Farrell, A. F. Hegarty, J. J. H. Miller, E. O’Riordan, G. I. Shishkin, *Robust Computational Techniques for Boundary Layers*, Chapman & Hall/CRC Press, Boca Raton, FL, 2000

Numerical solution of high accuracy for quasilinear singularly perturbed convection-diffusion problems

Lixuezheng

(School of Economics and Management, North China Univ. of Tech., Beijing, 100144)

Abstract In this paper, we mainly study to solve quasilinear singularly perturbed convection diffusion boundary value problems from two parts as follows:

(1) The first part, a class of hybrid difference schemes with variable weights on Bakhvalov-Shishkin mesh is proposed to compute the solution in quasilinear singularly perturbed convection – diffusion boundary value problems. First, we use hybrid difference schemes on Bakhvalov-Shishkin to study the general situation. We use the simple upwind scheme on the coarse part and central scheme on the fine part. Then, we define the discrete linear operator, and prove the discrete linear operator is an M-matrix. We get the uniform second-order $O(N^{-2})$ using the comparison principle. Moreover, we improved the variable weights. Finally, numerical example support our error estimates.

(2) In the second part, a class of hybrid difference schemes with variable weights on Bakhvalov-Shishkin mesh is proposed to compute the derivative in quasilinear singularly perturbed convection – diffusion boundary value problems. We use backward difference at $x_{i-\frac{1}{2}}$, and get the uniform second-order $O(N^{-2})$ according to the solution’s proof method. Then, we improve the previous papers in S mesh. Only if N is large enough, we have the hybrid scheme on S mesh and obtain uniform conver-

gence of nearly second-order $O(N^{-2} \ln^2 N)$. Last, we compare the S mesh and B-S mesh, and the error estimates of numerical solutions and numerical derivatives are all confirmed through our numerical experiments.

Key words quasilinear singularly perturbed convection-diffusion; layer adapted meshes; hybrid schemes, numerical derivatives; uniform convergence

几类线性模型中的 Bootstrap 方法及其应用

研究生 梅 波

导师 徐礼文

(北方工业大学理学院 北京 100144)

摘 要 本文研究了几类线性模型当中的假设检验问题,主要包括均值相等性检验和线性回归模型中参数显著性检验。本文分别提出了三种 bootstrap 检验方法解决上述三类线性模型的检验问题,首先将 bootstrap 方法与得分检验结合,给出一种解决 Behrens-Fisher 问题的方法,其次基于极大似然估计的思想,给出解决多个未知异方差总体的均值检验方法,最后在二次型检验统计量的基础上提出参数 bootstrap 检验方法。通过 Monte Carlo 模拟,在解决 Behrens-Fisher 问题 and 多正态总体的异方差均值检验问题中,所提出的参数 bootstrap 检验在控制第一类错误和检验势函数两方面都要优于传统的近似 t 检验和广义 F 检验,而且在样本量较小的情况下,检验效果均令人满意。针对 CAPM 高维情形下的有效性检验,本文结合广义加号逆的性质提出一种参数 bootstrap 检验方法,该方法可以用于样本维度大于样本期数的情况,应用范围更加广泛。

关 键 词 Behrens-Fisher 问题; Bootstrap 重抽样; 广义 p 值; CAPM 有效性

作者简介 梅波,男,1991 年出生,硕士研究生,主要研究方向为混合效应模型。

E-mail: meibo119@126.com

1 绪论

在两个总体的均值比较问题中,当两总体方差任意且未知时,称该问题为 Behrens-Fisher 问题^[3,4]。Welch^[5,6]给出近似 t 分布检验,此方法操作简单且较为精确。但此方法建立在随机自由度近似 t 分布基础上,自由度依赖于样本方差和样本容量,从而在一定情况下会出现误差。金华,郑圣听和陈伟权^[7]提出了一种与极大似然思想相结合的检验方法,即得分(Score)检验。检验的操作只依赖于标准正态分布表,方便操作,并且在控制第一类错误能力方面要优于 Welch 的近似 t 检验。

针对多个总体的异方差情况下的均值检验问题,目前也已有多种不同程度的近似检验方法,研究表明,Welch(1951)^[11]检验和 Weerahandi(1995)^[12]的广义 F(GF)检验在样本量适中或较大时才能在控制第一类错误方面表现良好。在 Dajani(2002)^[13]的数值模拟结果中发现,在比较总体的数目较多和样本量较小时,上述检验都不能很好的控制犯第一类错误的概率。Krishnamoorthy, Lu 和 Mathew (2007)^[14]基于组间加权平方和提出了一种参数 bootstrap 检验,并适用于固定效应和随机效应的检验。

在线性回归模型当中回归参数的检验也是研究重点之一。其中比较引人注目的是资本资产定价模型(CAPM)的检验。现有的大部分资产组合有效性检验方法都是建立在二次型 $\hat{\alpha}'V\hat{\alpha}$ 上的,其中 $\hat{\alpha}$ 是 α 的最小二乘估计, V 是一个正定矩阵。然而当 $T < N$ 时,样本协方差矩阵 $\hat{\Sigma}$ 退化成为非正定矩阵,从而导致无法求得矩阵的逆。为克服这个困难,Pesaran 和 Yamagata (2012)^[22]提出,一定程度上忽略样

本的相关性,选择 $V = \text{diag}(\hat{\Sigma})^{-1}$ 来避免求逆问题,他们得到二次型的标准化形式 $(W - EW)/\sqrt{\text{var}(W)}$,并且证明这个近似服从标准正态分布。

2 Behrens-Fisher 问题的参数 Bootstrap 检验

2.1 Welch 近似 t 检验

Welch^[4]指出,在 $H_0: \mu_1 = \mu_2$ 成立时,近似地有

$$T(\bar{X}_1, \bar{X}_2; S_1^2, S_2^2) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(df) \quad (2.3)$$

其中自由度为 $df = \frac{(S_1^2/n_1 + S_2^2/n_2)^2}{S_1^4/(n_1^2(n_1-1)) + S_2^4/(n_2^2(n_2-1))}$ 。

对于给定的显著性水平 α 和统计量 $(\bar{X}_1, \bar{X}_2; S_1^2, S_2^2)$ 的一个观测值 $(\bar{x}_1, \bar{x}_2; s_1^2, s_2^2)$,该检验的法则为:当 p 值 $P\{|t(df)| > |T(\bar{x}_1, \bar{x}_2; s_1^2, s_2^2)|\} < \alpha$ 时,拒绝原假设。

2.2 Score 检验

Score 检验统计量 U 近似地有:

$$U(\bar{X}_1, \bar{X}_2; S_1^2, S_2^2) = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \sim N(0, 1) \quad (2.4)$$

Score 检验法则为:对于给定的显著性水平 α ,当 $|u| > u_{\alpha/2}$ 时,拒绝原假设,这里 u 是 U 的一个观测值, $u_{\alpha/2}$ 表示标准正态分布的上侧 $\alpha/2$ 分位数点。

2.3 GP 检验

Tsui 和 Weerahandi^[23]通过拓展检验统计量的概念,给出如下广义检验变量 (Generalized Test Variable, GV),

$$GV = (X_1 - \bar{X}_2) \left(\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)^{-1/2} \times \left[\frac{\sigma_1^2 v_1^2}{n_1 V_1^2} + \frac{\sigma_2^2 v_2^2}{n_2 V_2^2} \right]^{1/2} \quad (2.5)$$

广义 p 值定义如下:

$$p = P_{H_0} \{ |GV| > |gv| \} = P \left\{ \left| T(n_1 + n_2 - 2)^{-1/2} \left(\frac{s_1^2}{B} + \frac{s_2^2}{1-B} \right)^{1/2} \right| \geq |\bar{x}_1 - \bar{x}_2| \right\} \quad (2.6)$$

GP 检验法则为:对于给定的显著性水平 α ,当 (2.5) 式中的 $p < \alpha$ 值时,拒绝原假设。

2.4 PB-t 检验

基于统计量 $T(\bar{X}_1, \bar{X}_2; S_1^2, S_2^2)$ 的 PB-t 检验枢轴量为:

$$T_B = \frac{\bar{X}_{B1} - \bar{X}_{B2}}{\sqrt{\frac{S_{B1}^2}{n_1} + \frac{S_{B2}^2}{n_2}}} \quad (2.7)$$

PB-t 检验的法则为:对于给定的显著性水平 α ,当 p 值 $P\{|T_B| > |T(\bar{x}_1, \bar{x}_2; s_1^2, s_2^2)|\} < \alpha$ 时,拒绝原假设。

2.5 PB-S 检验

类似 2.1.4 想法,基于式 (2.4) 中统计量 $U = (\bar{X}_1 - \bar{X}_2) / \sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}$ 的 PB-S 检验枢轴量为:

$$U_B = \frac{\bar{X}_{B1} - \bar{X}_{B2}}{\sqrt{\sigma_{B1}^2/n_1 + \sigma_{B2}^2/n_2}} \quad (2.8)$$

这里, $\sigma_{B1}^2, \sigma_{B2}^2$ 是在给定 \bar{X}_{B1}, S_{B1}^2 下通过条件极大似然估计得到的条件似然估计量, 其不依赖冗余参数。PB-S 检验的法则为: 对于给定的显著性水平 α , 当 p 值

$$P\{|U_B| > |U(\bar{x}_1, \bar{x}_2; s_1^2, s_2^2)|\} < \alpha \quad (2.9)$$

时, 拒绝原假设 H_0 , 这里 $U(\bar{x}_1, \bar{x}_2; s_1^2, s_2^2)$ 是 $U(\bar{X}_1, \bar{X}_2; S_1^2, S_2^2)$ 的观测值。

2.6. Monte Carlo 模拟研究

表 1 第一类错误结果显示, 在小样本下样本量相等时, Welch 近似 t 检验、PB-t 检验和 PB-S 检验控制第一类错误基本稳定在名义水平 0.05 上, 受方差比率的影响不是很明显。表 2 检验势函数的结果表明, 在两组样本量不相等时, Welch-t 检验和 PB-t 检验的势时而表现突出, 但同时不可避免它们犯第一类错误概率也偏大; Score 检验和 GP 检验受总体方差比率影响更加明显, 二者控制第一类错误都偏向保守, 检验势也劣与其他三种检验。PB-S 检验具有稳定且较好的势。

表 1 第一类错误概率的 Monte Carlo 估计值

样本容量	n=(5,5)					n=(4,8)					n=(8,8)					n=(10,15)				
总体方差	Welch-t	Score	GP	PB-t	PB-S	Welch-t	Score	GP	PB-t	PB-S	Welch-t	Score	GP	PB-t	PB-S	Welch-t	Score	GP	PB-t	PB-S
v(1, 0.01)	0.05	0.02	0.06	0.05	0.05	0.05	0.01	0.05	0.05	0.04	0.05	0.04	0.05	0.05	0.05	0.05	0.03	0.05	0.05	0.05
v(1, 0.1)	0.05	0.03	0.04	0.05	0.04	0.06	0.02	0.04	0.07	0.03	0.06	0.04	0.04	0.06	0.05	0.05	0.05	0.04	0.05	0.04
v(1, 0.3)	0.05	0.03	0.03	0.05	0.05	0.06	0.03	0.04	0.07	0.04	0.05	0.04	0.04	0.05	0.05	0.06	0.05	0.04	0.04	0.05
v(1, 0.5)	0.04	0.04	0.02	0.06	0.06	0.06	0.04	0.04	0.07	0.04	0.05	0.05	0.04	0.06	0.05	0.05	0.05	0.04	0.05	0.05
v(1, 0.7)	0.05	0.03	0.03	0.05	0.06	0.06	0.03	0.03	0.05	0.05	0.05	0.05	0.03	0.04	0.05	0.05	0.05	0.04	0.05	0.05
v(1, 0.9)	0.04	0.03	0.02	0.05	0.05	0.05	0.03	0.03	0.05	0.04	0.05	0.05	0.03	0.05	0.05	0.05	0.05	0.04	0.04	0.04
v(1, 1.1)	0.04	0.03	0.02	0.05	0.05	0.05	0.04	0.03	0.05	0.04	0.05	0.05	0.04	0.05	0.05	0.05	0.05	0.04	0.05	0.05
v(1, 1.3)	0.05	0.03	0.02	0.04	0.05	0.05	0.04	0.03	0.05	0.05	0.05	0.05	0.03	0.05	0.05	0.05	0.05	0.04	0.06	0.06
v(1, 1.5)	0.05	0.03	0.03	0.04	0.04	0.05	0.03	0.03	0.05	0.06	0.05	0.05	0.03	0.04	0.06	0.05	0.05	0.04	0.05	0.06
v(1, 1.7)	0.05	0.03	0.03	0.04	0.06	0.05	0.03	0.03	0.05	0.05	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.03	0.05	0.05
v(1, 1.8)	0.05	0.03	0.02	0.05	0.04	0.04	0.03	0.03	0.05	0.06	0.05	0.05	0.03	0.05	0.05	0.05	0.05	0.04	0.05	0.05

表 2 检验函数势的 Monte Carlo 模拟估计值

样本量		n=(8,8)						n=(5,10)					
总体方差	检验方法	m(0,0)	m(0,0.3)	m(0,0.6)	m(0,1)	m(0,1.5)	m(0,2)	m(0,0)	m(0,0.3)	m(0,0.6)	m(0,1)	m(0,1.5)	m(0,2)
v(1, 0.01)	Welch-t	0.05	0.12	0.30	0.67	0.95	1.00	0.05	0.08	0.18	0.39	0.71	0.90
	Score	0.04	0.09	0.26	0.62	0.92	1.00	0.03	0.04	0.09	0.24	0.50	0.76
	GP	0.05	0.11	0.31	0.67	0.95	1.00	0.04	0.07	0.17	0.38	0.70	0.90
	PB-t	0.05	0.11	0.32	0.68	0.95	1.00	0.05	0.08	0.18	0.39	0.72	0.90
	PB-S	0.05	0.11	0.31	0.66	0.95	1.00	0.05	0.07	0.17	0.40	0.71	0.90
v(1, 0.1)	Welch-t	0.05	0.11	0.30	0.65	0.94	1.00	0.05	0.09	0.19	0.40	0.71	0.90
	Score	0.04	0.10	0.26	0.62	0.92	1.00	0.03	0.04	0.11	0.25	0.51	0.75
	GP	0.05	0.10	0.29	0.64	0.93	1.00	0.04	0.08	0.17	0.37	0.68	0.90
	PB-t	0.05	0.12	0.31	0.66	0.94	1.00	0.06	0.10	0.19	0.42	0.72	0.91
	PB-S	0.05	0.11	0.30	0.65	0.94	1.00	0.05	0.07	0.15	0.36	0.67	0.89

样本量		n=(8,8)						n=(5,10)					
v(1, 0.6)	Welch-t	0.05	0.09	0.24	0.53	0.87	0.98	0.05	0.08	0.18	0.39	0.70	0.90
	Score	0.05	0.10	0.23	0.52	0.86	0.98	0.04	0.06	0.14	0.30	0.57	0.79
	GP	0.03	0.07	0.18	0.49	0.83	0.97	0.03	0.06	0.13	0.31	0.62	0.86
	PB-t	0.05	0.09	0.23	0.54	0.86	0.99	0.06	0.08	0.19	0.41	0.70	0.90
	PB-S	0.05	0.09	0.25	0.54	0.87	0.99	0.04	0.08	0.15	0.36	0.64	0.85
v(1, 1.2)	Welch-t	0.05	0.08	0.19	0.42	0.75	0.94	0.05	0.08	0.15	0.35	0.64	0.86
	Score	0.05	0.08	0.19	0.41	0.74	0.94	0.04	0.07	0.14	0.30	0.58	0.78
	GP	0.03	0.07	0.14	0.36	0.71	0.91	0.03	0.05	0.13	0.25	0.53	0.81
	PB-t	0.04	0.07	0.19	0.40	0.74	0.94	0.06	0.08	0.16	0.37	0.66	0.87
	PB-S	0.05	0.09	0.20	0.44	0.76	0.94	0.05	0.08	0.15	0.34	0.60	0.83
v(1, 1.8)	Welch-t	0.05	0.07	0.15	0.33	0.64	0.87	0.05	0.07	0.14	0.30	0.59	0.83
	Score	0.05	0.07	0.15	0.33	0.64	0.86	0.04	0.06	0.13	0.29	0.54	0.78
	GP	0.03	0.05	0.12	0.29	0.58	0.84	0.03	0.04	0.10	0.23	0.49	0.75
	PB-t	0.05	0.08	0.15	0.34	0.65	0.87	0.05	0.07	0.15	0.30	0.57	0.83
	PB-S	0.05	0.08	0.16	0.33	0.64	0.86	0.05	0.07	0.14	0.33	0.57	0.82

3 异方差下多总体均值比较的参数 bootstrap 检验

3.1. 极大似然 PB 检验

PB 检验变量如下:

$$T_B = \sum_{i=1}^n \frac{n_i}{\hat{\sigma}_{B_i}^2} \frac{(\bar{X}_{B_i} - \sum_{i=1}^n \hat{w}_{B_i} \bar{X}_{B_i})^2}{1 - \hat{w}_{B_i}} = (X_B^T \hat{\Delta}_B^{-1/2}) \hat{A}_B \hat{B}_B^{-1} \hat{A}_B (\hat{\Delta}_B^{-1/2} \bar{X}_B) \quad (3.13)$$

对于给定的一组样本观察值 $(\bar{x}_1, \bar{x}_2, \dots, \bar{x}_k; s_1^2, s_2^2, \dots, s_k^2)$ 和检验显著性水平 α , 当 $P\{T_B > T_0\} < \alpha$ 成立时, 就拒绝原假设 H_0 , 其中 T_0 是的一次观察值。

3.2. Welch 检验

记 $\omega_i = (n_i/S_i^2) / \sum_{i=1}^k (n_i/S_i^2)$, $i=1, 2, \dots, k$, 传统检验方法中用到的组间加权平方和统计量为:

$$T_N(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k; \sigma_1^2, \sigma_2^2, \dots, \sigma_k^2) = \sum_{i=1}^k \frac{n_i}{\sigma_i^2} (\bar{X}_i - \sum_{i=1}^k \omega_i \bar{X}_i)^2 \quad (3.14)$$

Welch(1951)给出了近似自由度的 F 检验统计量:

$$W^* = \frac{T_N(\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k; S_1^2, S_2^2, \dots, S_k^2)/(k-1)}{1 + (2(k-2)/(k^2-1)) \sum_{i=1}^k (1/(n_i-1))(1-\omega_i)^2} \stackrel{\text{近似}}{\sim} F_{k-1, f} \quad (3.15)$$

3.3 Monte Carlo 模拟研究

从表 3 得出本文提出的参数 bootstrap 检验在各种模拟参数设置组合下都能很好的将第一类错误控制在名义水平 0.05 附近, 相比传统的检验方法, PB 检验结果更加准确。从表 4 看出极大似然参数 bootstrap 检验在保证第一类错误的同时, 与传统的检验方法具有相似的检验功效, 保证了检验的有效性。

表 3 三种检验方法:第一类错误概率的 Monte Carlo 估计值

$k=3$	$n=(3,3,3)$			$n=(5,5,5)$			$n=(10,10,10)$			$n=(2,4,8)$			$n=(3,6,9)$		
$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	WT	GF	PB	WT	GF	PB	WT	GF	PB	WT	GF	PB	WT	GF	PB
(1,1,1)	0.04	0.02	0.05	0.05	0.03	0.05	0.05	0.05	0.06	0.08	0.04	0.07	0.06	0.05	0.05
(1,1,0.5)	0.04	0.03	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.08	0.05	0.06	0.07	0.05	0.05
(1,0.5,0.5)	0.04	0.03	0.05	0.04	0.03	0.04	0.05	0.05	0.05	0.10	0.06	0.06	0.06	0.05	0.04
(1,0.1,0.1)	0.06	0.05	0.05	0.04	0.04	0.05	0.06	0.05	0.04	0.11	0.06	0.06	0.07	0.06	0.06
(0.1,0.1,0.1)	0.04	0.03	0.05	0.04	0.04	0.05	0.05	0.05	0.05	0.08	0.04	0.06	0.05	0.05	0.05
(0.5,1,1)	0.04	0.03	0.06	0.04	0.04	0.05	0.05	0.05	0.05	0.06	0.04	0.06	0.06	0.04	0.06
(0.05,1,1)	0.06	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.04	0.05	0.03	0.05	0.05	0.04	0.05
(0.5,0.5,1)	0.04	0.03	0.05	0.04	0.04	0.05	0.05	0.04	0.05	0.07	0.03	0.06	0.06	0.04	0.05
$k=6$	a			b			c			d			e		
$(\sigma_1^2, \sigma_2^2, \dots, \sigma_6^2)$	WT	GF	PB	WT	GF	PB	WT	GF	PB	WT	GF	PB	WT	GF	PB
(1,1,1,1,1,1)	0.07	0.09	0.04	0.06	0.08	0.06	0.06	0.07	0.05	0.14	0.11	0.04	0.09	0.09	0.05
(1,0.8,0.6,0.4,0.2,0.1)	0.08	0.12	0.04	0.06	0.08	0.05	0.05	0.06	0.05	0.14	0.13	0.06	0.09	0.09	0.05
(0.5,0.5,0.5,0.5,0.5,0.5)	0.08	0.10	0.05	0.05	0.07	0.05	0.05	0.07	0.05	0.13	0.11	0.05	0.09	0.08	0.05
(0.5,0.3,0.1,0.05,0.03,0.01)	0.10	0.12	0.06	0.07	0.09	0.05	0.06	0.07	0.05	0.16	0.13	0.05	0.09	0.09	0.06
(1,0.3,0.8,0.01,0.05,0.5)	0.11	0.14	0.05	0.07	0.10	0.05	0.06	0.07	0.05	0.15	0.12	0.06	0.08	0.09	0.05
(0.2,0.9,1,0.1,0.04,0.5)	0.09	0.13	0.05	0.07	0.09	0.05	0.06	0.06	0.05	0.15	0.12	0.05	0.09	0.08	0.05
(0.5,0.05,0.01,0.8,0.3,1)	0.11	0.14	0.04	0.09	0.09	0.05	0.06	0.07	0.04	0.16	0.11	0.05	0.09	0.08	0.05
(0.5,0.04,0.1,1,0.9,0.2)	0.10	0.12	0.04	0.07	0.10	0.05	0.05	0.07	0.05	0.15	0.11	0.05	0.08	0.08	0.06
$k=10$	f			g			h			i			j		
$(\sigma_1^2, \sigma_2^2, \dots, \sigma_{10}^2)$	WT	GF	PB	WT	GF	PB	WT	GF	PB	WT	GF	PB	WT	GF	PB
(1,1,1,1,1,1,1,1,1,1)	0.14	0.23	0.04	0.07	0.14	0.05	0.06	0.09	0.06	0.17	0.15	0.05	0.10	0.10	0.05
(1,0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1)	0.16	0.24	0.05	0.09	0.14	0.04	0.06	0.09	0.05	0.18	0.15	0.05	0.10	0.11	0.06
(1,0.8,0.6,4,0.2,0.1,0.08,0.06,0.04,0.02)	0.17	0.25	0.05	0.09	0.15	0.05	0.06	0.09	0.05	0.18	0.14	0.05	0.10	0.09	0.06
(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5)	0.14	0.23	0.04	0.09	0.14	0.05	0.06	0.08	0.05	0.15	0.15	0.04	0.09	0.10	0.05
(0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1)	0.15	0.24	0.05	0.09	0.15	0.06	0.06	0.09	0.04	0.13	0.12	0.06	0.08	0.08	0.05
(0.02,0.04,0.06,0.08,0.1,0.2,4,0.6,0.8,1)	0.18	0.25	0.04	0.09	0.15	0.06	0.06	0.09	0.05	0.09	0.09	0.04	0.06	0.08	0.05
(0.8,0.9,0.09,0.8,1,0.2,0.3,0.9,0.3,0.6)	0.14	0.22	0.05	0.09	0.14	0.05	0.05	0.09	0.05	0.17	0.13	0.06	0.09	0.09	0.04
(1,0.4,1,0.9,0.9,0.1,0.2,0.1,1,0.03)	0.15	0.25	0.04	0.08	0.15	0.04	0.06	0.08	0.05	0.18	0.14	0.05	0.11	0.12	0.05
$k=20$										k			L		
$(\sigma_1^2, \sigma_2^2, \dots, \sigma_{20}^2)$										WT	GF	PB	WT	GF	PB
(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)										0.28	0.52	0.04	0.17	0.16	0.05
(1,1,0.9,0.9,0.8,0.8,0.7,0.7,0.6,0.6,0.5,0.5,0.4,0.4,0.3,0.3,0.2,0.2,0.1,0.1)										0.30	0.52	0.05	0.19	0.16	0.05
(1,0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1,0.09,0.08,0.07,0.06,0.05,0.04,0.03,0.02,0.01,0.01)										0.32	0.52	0.04	0.20	0.16	0.06
(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5)										0.29	0.50	0.04	0.18	0.15	0.05
(0.8,0.8,0.8,0.5,0.5,0.5,0.4,0.4,0.4,0.1,0.1,0.1,0.05,0.05,0.05,0.05,0.01,0.01,0.01,0.01)										0.31	0.52	0.05	0.19	0.15	0.05

$k=20$	k			L		
$(\sigma_1^2, \sigma_2^2, \cdots, \sigma_{20}^2)$	WT	GF	PB	WT	GF	PB
$(0.01, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)$	0.31	0.54	0.04	0.12	0.12	0.04
$(0.01, 0.01, 0.01, 0.01, 0.05, 0.05, 0.05, 0.05, 0.1, 0.1, 0.1, 0.4, 0.4, 0.4, 0.5, 0.5, 0.5, 0.8, 0.8, 0.8)$	0.31	0.53	0.04	0.14	0.14	0.05

注:表中 $a. n=(3,3,3,3,3,3); b. n=(5,5,5,5,5,5); c. n=(10,10,10,10,10,10); d. n=(2,4,6,8,15,28); e. n=(3,6,9,15,32,35); f. n=(3,3,3,3,3,3,3,3,3,3); g. n=(5,5,5,5,5,5,5,5,5,5); h. n=(10,10,10,10,10,10,10,10,10,10); i. n=(2,4,6,8,10,15,20,30,35,40); j. n=(3,6,9,12,15,20,30,35,40,45); k. n=(3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3); l. n=(2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40)$

表 4 三种检验方法:检验函数势的 Monte Carlo 模拟估计值

$k=3, \mu=0$							
$n=(10,10,10)$		(μ_2, μ_3)					
$(\sigma_1^2, \sigma_2^2, \sigma_3^2)$	Tests	(0,0)	(0,0.3)	(0,0.6)	(0,0.9)	(0.5,1)	(1.5,1)
(1,0.3,0.9)	WT	0.05	0.10	0.29	0.57	0.44	0.93
	GF	0.04	0.09	0.27	0.55	0.45	0.92
	PB	0.06	0.09	0.21	0.43	0.28	0.87
(1,0.1,0.5)	WT	0.04	0.15	0.50	0.84	0.58	0.98
	GF	0.05	0.16	0.47	0.85	0.61	0.98
	PB	0.05	0.15	0.41	0.69	0.39	0.94
$k=10, (\mu_1, \cdots, \mu_8)=0$							
$n=(15,15,15,20,20,20,25,25,25,25)$		(μ_9, μ_{10})					
$(\sigma_1^2, \sigma_2^2, \cdots, \sigma_{10}^2)$	Tests	(0,0)	(0,0.3)	(0,0.6)	(0,0.9)	(0.5,1)	(1.5,1)
a	WT	0.06	0.12	0.48	0.86	0.98	1.00
	GF	0.06	0.17	0.51	0.90	0.99	1.00
	PB	0.05	0.13	0.37	0.70	0.95	1.00
b	WT	0.05	0.17	0.60	0.95	1.00	1.00
	GF	0.07	0.19	0.63	0.95	1.00	1.00
	PB	0.04	0.14	0.45	0.83	0.99	1.00

注: $a. =(1,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9); b. =(1,0.1,0.1,0.1,0.3,0.3,0.3,0.7,0.7,0.7)$

3.4. 两种 PB 检验的比较

本节进一步讨论我们提出的 PB 检验与 Krishnamoorthy, Lu 和 Mathew^[14]提出的 PB 检验的异同点,旨在为 bootstrap 方法与其他的统计方法相结合提供一种思路。此处为了区别两种 PB 检验,本节用表示本文的极大似然参数 bootstrap 检验,表示 Krishnamoorthy, Lu 和 Mathew^[14]提出的参数 bootstrap 检验方法。表 5 是选取总体数分别为 6 和 20 的模拟结果。从整体上看,当样本量较大时二者都能比较好的控制第一类错误,但是当样本量较小时检验控制第一类错误的能力有明显的不足。相比之下,检验在各种参数设置上都能较好的控制第一类错误。

表 5 两种 PB 检验:第一类错误概率的 Monte Carlo 估计值

$k=6$	a		b		c		d		e		f	
$(\sigma_1^2, \sigma_2^2, \dots, \sigma_6^2)$	PB _K	PB _{ML}	PB _K	PB _{ML}	PB _K	PB _{ML}	PB _K	PB _{ML}	PB _K	PB _{ML}	PB _K	PB _{ML}
(1, 1, 1, 1, 1, 1)	0.03	0.05	0.05	0.05	0.05	0.06	0.12	0.05	0.07	0.06	0.05	0.05
(1, 0.5, 0.25, 0.1, 0.05, 0.02)	0.04	0.06	0.05	0.05	0.05	0.05	0.10	0.06	0.05	0.04	0.05	0.05
(0.5, 0.3, 0.1, 0.05, 0.03, 0.01)	0.05	0.04	0.05	0.04	0.05	0.06	0.11	0.05	0.05	0.05	0.04	0.05
(0.1, 0.2, 0.4, 0.6, 0.8, 1)	0.04	0.04	0.05	0.05	0.05	0.05	0.07	0.06	0.05	0.04	0.05	0.05
(0.02, 0.05, 0.1, 0.25, 0.5, 1)	0.05	0.04	0.05	0.05	0.04	0.04	0.06	0.05	0.06	0.06	0.04	0.05
(0.5, 0.04, 0.1, 1, 0.9, 0.2)	0.03	0.04	0.05	0.05	0.05	0.05	0.11	0.05	0.06	0.05	0.05	0.05
$k=20$					g		h		i		j	
$(\sigma_1^2, \sigma_2^2, \dots, \sigma_{20}^2)$					PB _K	PB _{ML}	PB _K	PB _{ML}	PB _K	PB _{ML}	PB _K	PB _{ML}
(1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)					0.02	0.04	0.05	0.05	0.11	0.05	0.05	0.05
(1,0.9,0.8,0.7,0.6,0.5,0.4,0.3,0.2,0.1,0.09,0.08,0.07,0.06,0.05,0.04,0.03,0.02,0.01,0.01)					0.03	0.05	0.05	0.06	0.08	0.04	0.05	0.06
(0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5,0.5)					0.02	0.04	0.05	0.05	0.11	0.04	0.04	0.05
(0.8,0.8,0.8,0.5,0.5,0.5,0.4,0.4,0.4,0.1,0.1,0.1,0.05,0.05,0.05,0.05,0.01,0.01,0.01,0.01)					0.04	0.05	0.04	0.05	0.08	0.04	0.05	0.05
(0.01,0.01,0.02,0.03,0.04,0.05,0.06,0.07,0.08,0.09,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1)					0.03	0.05	0.05	0.06	0.08	0.05	0.04	0.05
(0.01,0.01,0.01,0.01,0.05,0.05,0.05,0.05,0.1,0.1,0.1,0.4,0.4,0.4,0.5,0.5,0.5,0.8,0.8,0.8)					0.04	0.05	0.05	0.05	0.10	0.05	0.05	0.05

注:表中 $a. n=(3,3,3,3,3,3); b. n=(5,5,5,5,5,5); c. n=(10,10,10,10,10,10); d. n=(2,4,6,8,15,28); e. n=(3,6,9,15,32,35); f. n=(10,15,20,25,30,10); g. n=(3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3,3); h. n=(10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10,10); i. n=(2,4,6,8,10,12,14,16,18,20,22,24,26,28,30,32,34,36,38,40); j. n=(10,14,18,22,26,30,34,38,42,46,50,54,58,62,66,70,74,78,82,86)$

4 高维情况下 CAPM 的参数 bootstrap 检验

4.1 PY 近似检验

下面简单给出 PY 检验的检验统计量:

$$J_a = \frac{W_d - E(W_d)}{\sqrt{\text{Var}(W_d)}} = \frac{N^{-1/2} \sum_{i=1}^N \left(t_i^2 - \frac{\nu}{\nu-2} \right)}{\sqrt{\left(\frac{\nu}{\nu-2} \right)^2 \frac{2(\nu-1)}{\nu-4} [1 + (N-1)\bar{\rho}^2 + O(\nu^{-1/2})]}} \quad (4.5)$$

其中: $W_d = \sum_{i=1}^N \frac{\hat{\nu} \hat{\alpha}_i}{(1 + \hat{r}_m^2 / \hat{\sigma}_m^2) \hat{\sigma}_i^2}$, $\nu = T - 2$,

$\bar{\rho}^2 = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j=1}^{i-1} \rho_{ij}^2 I(\nu \rho_{ij}^2 > \theta_N)$, $\sqrt{\theta_N} = \Psi^{-1}(1 - p_N/2)$, $\Psi^{-1}(\cdot)$ 为标准正态分布分布函数的逆

函数, $I(\cdot)$ 为示性函数, $p_N = 1 - (1 - p)^{1/(N-1)}$ 表示 $N-1$ 个独立检验 $\rho_{ij} = 0, i \neq j$ 的显著性水平, 本文

在模拟当中取 $p=0.1$ 。在弱相关性下,当原假设成立时,近似有 $J_a \sim N(0,1)$ 。

4.2. SSN 检验和 WSN 检验

SSN 检验是建立在下面的符号检验统计量上:

$$SS_N = \sum_{i=1}^N S_i^2 \quad (4.6)$$

$$\text{其中 } S_i = \frac{\left[\sum_{t=1}^{\Gamma} I(z_{it} > 0) \right] - \frac{\Gamma}{2}}{\sqrt{\frac{\Gamma}{4}}}, z_{it} = \left(\frac{r_{i,t+\Gamma} - r_{i,t}}{r_{pt+\Gamma} - r_{pt}} \right) \left(\frac{r_{pt} - r_{pt+\Gamma}}{r_{pt} - r_{pt+\Gamma}} \right), t=1, 2, \dots, \Gamma, I(A) \text{ 是示性函数, } \Gamma$$

是 $T/2$ 的整数部分。

WSN 检验的检验统计量是:

$$WS_N = \sum_{i=1}^N W_i^2 \quad (4.7)$$

$$\text{其中 } W_i = \frac{\left[\sum_{t=1}^{\Gamma} I(z_{it} > 0) \text{Rank}(|z_{it}|) \right] - \Gamma(\Gamma+1)/4}{\sqrt{\Gamma(\Gamma+1)(2\Gamma+1)/24}}, \text{Rank}(|z_{it}|) \text{ 是 } |z_{it}| \text{ 的秩。}$$

Gungor 和 Luger (2009)^[26]证明了在原假设 $H_0: \alpha_i = 0$ 下,当 $\Gamma \rightarrow \infty$ 时 S_i, W_i 有极限近似标准正态分布,所以 SS_N, WS_N 都近似服从 χ_n^2 分布。

4.3 参数 Bootstrap 检验:PBW 检验

我们构造的 PB 检验变量形式是:

$$PBW \equiv \hat{\alpha}' \hat{\Sigma}^+ \hat{\alpha} \quad (4.8)$$

$$\text{其中 } \hat{\Sigma}^+ \text{ 为加号逆: } \hat{\Sigma}^+ = Q \begin{bmatrix} \Lambda^{-1} & 0 \\ 0 & 0 \end{bmatrix} P', Q, P \text{ 为正交矩阵, } \Lambda \text{ 是 } N \times N \text{ 的对角矩阵, 是对角}$$

$$\text{元素是矩阵 } \hat{\Sigma} \text{ 的奇异值分解所对应的矩阵 } \hat{\Sigma} = P \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix}。$$

4.4. Monte Carlo 模拟研究

本节对上面的提出的 PBW 检验,以及 PY 检验、SSN 检验和 WSN 检验进行模拟,分别计算三种检验方法控制第一类错误和检验的势。利用新的参数设置方式,模拟 bootstrap 检验 PBW 的第一类错误,并且比较研究各种方法受相关性的影响。结果如表 7 从表中可以看出,在假设总体的协方差阵为奇异矩阵的情况下,即高维情况,各种检验方法受相关性强弱的影响要小的多,相比表 6 结果表现得不是很明显。而且只有 PBW 检验方法可以很好的控制第一类错误,而其他几种方法犯第一类错误的概率均高于名义水平。

表 7 调整协方差阵下不同强弱相关性下的第一类错误 ($N=48, T=24$)

Test		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
PY	1%	0.060	0.066	0.064	0.065	0.071	0.072	0.066	0.067	0.066	0.058
	5%	0.132	0.150	0.150	0.148	0.150	0.145	0.146	0.130	0.126	0.114
	10%	0.208	0.240	0.240	0.240	0.221	0.218	0.219	0.214	0.203	0.176
SSN	1%	0.040	0.041	0.042	0.039	0.050	0.056	0.055	0.066	0.067	0.074
	5%	0.121	0.106	0.102	0.112	0.110	0.113	0.123	0.118	0.123	0.145

Test		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	10%	0.173	0.162	0.164	0.167	0.164	0.167	0.179	0.175	0.174	0.194
WSN	1%	0.037	0.047	0.043	0.057	0.071	0.064	0.062	0.070	0.068	0.084
	5%	0.103	0.119	0.104	0.122	0.130	0.125	0.128	0.146	0.143	0.153
	10%	0.158	0.180	0.165	0.190	0.184	0.177	0.183	0.204	0.194	0.203
PB	1%	0.012	0.012	0.008	0.012	0.018	0.014	0.008	0.020	0.01	0.013
	5%	0.048	0.046	0.040	0.052	0.058	0.054	0.048	0.042	0.05	0.047
	10%	0.090	0.096	0.094	0.090	0.114	0.090	0.106	0.104	0.09	0.101

5 总结与建议

本文结合 bootstrap 方法的优点,与传统的方法相结合,避免了传统方法的一些不足之处,同时也增强了方法的适用性。文中提出的参数 bootstrap 均值检验,在总体方差未知且不相等的情况下,可以有效地检验均值是否相等,并且在样本量较小时,参数 bootstrap 检验在控制第一类错误和势方面都优于传统的 welch t 检验和广义 F 检验。此外针对 $N > T$ 情况下,传统 CAPM 有效性的检验方法遇到无法逾越的障碍,在调整检验统计量并结合 bootstrap 方法下,提出的 PBW 检验可以很好的满足检验的要求。并且通过与已有的针对相应的高维问题检验方法相比较,可以发现 bootstrap 方法更贴近实际情况,检验的效果也是令人满意的。

参 考 文 献

- [1] B Efron. Bootstrap methods: Another look at the Jackknife. *Ann Statistics*. 1979b(7): 1—26
- [2] B Efron, R. J Tibshirani. *An Introduction to Bootstrap*. Chapman&Hall. Lonon, 1993
- [3] B V Behrens. Ein beitrage zur fehlerberechnung bei wenige beobachtungen. *Landwirtsch. Jb.* 1929(6):807—837
- [4] R A Fisher. The fiducial argument in statistical inference [J]. *Annals of Eugenics*, 1935(6):391 - 398
- [5] B L Welch. The Specification of Rules for Rejecting Too Variable a Product, with Particular Reference to an Electric Lamp Problem [J]. *Supplement to the Journal of the Royal Statistical Society*. 1936(3):29—48
- [6] B L Welch. The Significance of the Difference Between Two Means when the Population Variances are Unequal [J]. *Biometrika*, 1938(29):350—362
- [7] 金华, 郑圣听, 陈伟权. Behrens-Fisher 问题的正态逼近[J]. *统计研究*, 2009(26):106—108
- [8] G S James, The comparison of several groups observations when the ratios of population variances are unknown. *Biometrika*, 1951(38): 324—329
- [9] D J Bishop, E J Dudewicz. Exact analysis of variance with unequal variances: test procedures and tables. *Technometrics*, 1978(20):419—430
- [10] S Lee, C H Ann. Modified ANOVA for unequal variances [J]. *Comm. Statist. Simulation Comput.* 2003(32): 987—1004
- [11] B L Welch. On the comparison of several mean values:an alternative approach[J]. *Biometrika*, 1951(38):330—336
- [12] S Weerahandi. ANOVA under unequal error variances [J]. *Biometrika*. 1995(51):589—599

- [13] A N Dajani. Contributions to statistical inference for some fixed and random models [M], PhD. Dissertation, Department of Mathematics and Statistics, University of Maryland, Baltimore County
- [14] K Krishnamoorthy, F Lu, T Mathew. A parametric bootstrap approach for ANOVA with unequal variances: Fixed and random models [J]. Computational Statistics & Data Analysis. 2007(51):5731—5742
- [15] Gibbons M, Ross S, Shanken J. A test of the efficiency of given portfolio [J]. Econometrica. 1989(57): 1121 — 1152
- [16] J Shanken: A Bayesian approach to testing portfolio efficiency [J]. Journal of Financial Economics. 1987(19): 195 - 216
- [17] Harvey C. R. and G. Zhou. Bayesian Inference in Asset Pricing Tests [J]. Journal of Financial Economics. 1990 (26): 221—254
- [18] S Kandel, R McCulloch and R F Stambaugh: Bayesian inference and portfolio efficiency. Review of Financial Studies. 1995(8):1 - 53
- [19] G Zhou: Asset-pricing Tests under Alternative Distributions [J]. The Journal of Finance. 1993(47):1927—1942
- [20] A C MacKinlay and M P Richardson: Using Generalized Method of Moments to Test Mean-Variance Efficiency [J]. The Journal of Finance, 1991(46):511—527
- [21] P H Chou and H F Zhou: Using Bootstrap to Test Portfolio Efficiency [J]. Annals of Economics and Finance. 2006(2):217—249
- [22] M H Pesaran and T Yamagata: Testing CAPM with a large number of assets. Manuscript. 2012
- [23] K W Tsui, S. Weerahandi. Generalized p-Values in Significance Testing of Hypotheses in the Presence of Nuisance Parameters [J]. Journal of the American Statistical Association. 1989 (84):602—607
- [24] Sharpe W F. Capital asset prices: a theory of market equilibrium under conditions of risk [J]. The Journal of Finance. 1964(19): 425 — 442
- [25] Lintner J. The valuation of risk assets and the selection of risky investments in stock portfolio and capital budgets [J]. The Review of Economics and Statistics. 1965(47):13 — 37
- [26] S Gungor and R Luger: Exact distribution-free tests of mean-variance efficiency [J]. Journal of Empirical Finance. 2009(16): 816—829
- [27] D TU, J Shao. Jackknife Bootstrap. Springer Verlag, New York. 1995
- [28] 史树中. 金融经济学十讲 [M]. 上海: 上海人民出版社, 2011. 11
- [29] 黎子良, 邢海鹏. 金融市场中的统计模型和方法 [M]. 北京: 高等教育出版社, 2012. 11
- [30] 王松桂, 史建红等. 线性模型引论 [M]. 北京: 科学出版社, 2011. 1
- [31] 张尧庭, 方开泰. 多元统计分析引论 [M]. 北京: 科学出版社, 2003. 9

Analysis of Bootstrap methods and its applications in a few linear model classes

Mei Bo

(School of Economics and Management, North China Univ. of Tech., Beijing, 100144)

Abstract This paper focuses on analyzing problems in a few linear-model classes, mainly concerning about the test of equality of a group of means and significant test of estimated parameters in

linear regression model. Firstly, we study the equality of two normal means under different and unknown variances, called Behrens-Fisher problem. We combine bootstrap idea and the existed score test and get a better test method for Behrens-Fisher problem. Based on this work, we further consider the equality test of means in the case which involves more than two normal populations. Drawing on Max Likelihood idea, we derive a parametric bootstrap method to solve the problem of testing a groups of normal means under heteroscedasticity. Finally, we manage to incorporate bootstrap method into traditional quadratic test variable to avoid solving the inverse of singular covariance in testing the significance of estimated parameter in a multivariable linear model-CAPM, especially when the number of the excess returns we study is much bigger than sample size. Through Monte Carlo simulation, the parametric bootstrap tests we proposed perform better than traditional welch t-test and generalized F test in terms of type I error and power both in comparing two normal means and more than two ones, especially when the sizes of all groups is small and variant. As for the efficiency test of CAPM in high dimension case, the standard inverse of symmetric matrix is replaced by a special generalized counterpart, the Moore-Penrose Inverse and we proposed parametric bootstrap test, which could be applied to more general situations.

Key words behrens-fisher; bootstrap re-sampling; generalized p-value; CAPM model

基于价格条件 VaR 的套保套利研究

研究生 杨桂君

导师 肖春来

(北方工业大学理学院 北京 100144)

摘 要 沪深 300 股指期货作为我国推出的第一支股指期货,对我们证券市场具有很大意义。由于股指期货本身具有转移风险、价格发现、优化资源配置特点,我们利用这些特点,进行有效的套期保值和套利操作。本文就是对沪深 300 股指期货的套期保值和套利进行深入研究,第一,利用差分法和序列分解法对沪深 300 股票指数期货日收盘价的对数进行建模分析,预测日收盘价的走势,并依次评价这两种方法的预测效果;第二,利用最小二乘法和误差修正模型两种方法求解最优套期保值比率,根据求解的最优套保比率计算股指期货合约数量,并进行套期保值操作,并对实际套保效果进行评估;第三,根据价格条件 VaR 方法对套利模型风险进行测算,并寻找最佳套利时机。

关键词 股指期货;套期保值;套利;价格条件 VaR

作者简介 杨桂君,女,1990 年出生,硕士研究生,主要研究方向为经济统计。

E-mail: ygjketty4@163.com

1 绪论

美国堪萨斯期货交易所首次推出股指期货,它以自身的高杠杆、低交易成本的优点大大丰富了投资者的交易策略和交易工具,股指期货已经逐步成为证券市场发展新的衍生金融工具的着力点。我国的沪深 300 股指期货于 2010 年 4 月 16 日正式上市,这一期货类型的出现大大丰富了我国股市的交易结构,我国众多投资者有了更多的投资机会。

沪深 300 股指期货是本文主要的研究对象,本文将使用计量经济的方法,通过股指期货套期保值模型和套利模型对沪深 300 股指期货进行更深入的研究。

随着证券市场的越加多元化,投资需求越来越多,与此同时,风险控制成为重要的研究课题。William J. Baumol(1963)^[1]在“An Expected Gain-Confidence Limit Criterion for Portfolio Selection”首先提出了 VaR 思想;90 年代,J. P. 摩根集团将 VaR 思想运用到了 RiskMetrics 风险管理系统中。VaR 是一个全面衡量金融投资组合风险的单一参数,它已经被公司及金融机构广泛应用。VaR 的定义是:在市场正常情况下,在一定的置信水平,给定的一段时间内,某个金融资产的预期最大损失,表达式为: $P(X(\omega) \leq VaR) = \alpha$,其中: $X(\omega)$ 为一定时期内,某一金融资产或资产组合 ω 的损失值, α 为置信度。Darrell Duffie 和 Jun Pan(1997)^[2]总结了 VaR 各种模型以及 VaR 的各种计算方法,认为 VaR 是目前风险度量中最有效最可行的方法。国内最早引入 VaR 理论的是王春峰(2001)^[4]的著作《金融市场风险管理》,在书中阐述了 VaR 的定义和公式,以及计算方法和模型,并且将 VaR 应用到实际的金融市场风险控制中,这为我国的金融风险价值研究奠定了一定的基础。之后学者开始运用 VaR 方法展开实证研究,肖春来、宋然(2003)^[6]运用 VaR 理论对我国的深沪股市进行研究,认为 VaR 方法适用于我国股市风险管理。

肖春来、柴文义、扬威(2003)^[8]在“条件 VaR 理论的应用于研究”中认为传统 VaR 理论和 CVaR 理论的股票收益率的统计分布特征在一定时期内基本稳定这一重要假设存在缺陷,随着市场条件的变化,股票收益率的统计分布特征会发生变化,不一定是平稳的,因此提出了条件收益率的思想,简称为条件 VaR,该理论改进了传统 VaR 和 CVaR 理论的假设缺陷,为股票投资组合和投资风险控制提供了更准确的方法,并深入研究了价格条件 VaR 测算问题,分别在股票的价格和收益率服从二维正态分布和二维 t 分布的假设前提下进行研究,推导出股票的条件收益率的分布特征。

肖春来、柴文义、刘喜波(2007)^[9]在“基于二维正态分布的条件 VaR 研究”中得到价格条件 VaR:

$$VaR = (1 - \rho)(\mu - \ln P_t) - Z_\alpha \times \sigma \sqrt{(1 - \rho^2)},$$

杨桂君、肖春来、常京宏(2014)^[10]在“基于二维 t 分布的条件 VaR 研究”中利用多元 t 分布研究价格条件的收益率分布问题,与正态分布相比,t 分布较好地刻画了证券收益率分布的尖峰厚尾现象。

2 沪深 300 股指期货价格研究

2.1 利用差分方法建立模型

选取沪深 300 股指期货 IF1403 合约的日收盘价数据 P_t ,时间区域从 2013 年 7 月 22 日到 2014 年 3 月 5 日。我们利用 ARMA 模型理论对沪深 300 指数的日收盘价进行建模分析及预测,根据模型拟合结果表明 ARIMA(1,2,0)模型的参数估计值中 AR(1)的系数具有显著性,其模型的数学表达式为:

$$dd\ln P_t = -0.409dd\ln P_{t-1} + \epsilon_t \quad (1)$$

整理后得到:

$$\ln P_t = 1.591\ln P_{t-1} - 0.182\ln P_{t-2} - 0.409\ln P_{t-3} + \epsilon_t \quad (2)$$

$$\text{收益率: } R_t = \ln P_t - \ln P_{t-1} = 0.591\ln P_{t-1} - 0.182\ln P_{t-2} - 0.409\ln P_{t-3} + \epsilon_t \quad (3)$$

2.2 利用序列分解方法建立模型

根据时间序列加法分解原理,我们将 $\ln P_t$ 序列分解为移动平均项和围绕移动平均的波动项。具体分解原理如下方式,令:

$$\ln P_t = ma_{t-1} + d_t \quad (4)$$

其中: $ma_{t-1} = \frac{1}{m} \sum_{k=1}^m \ln P_{t-k}$, m 为移动平均项数。

依据式(4),新序列 d_t 的分解公式如下:

$$d_t = \ln P_t - ma_{t-1} \quad (5)$$

依据式(5),对于 t 期的 $\ln P_t$ 、 d_t 来讲, ma_{t-1} 是已知的,那么对于 $\ln P_t$ 的数据建模就变成了对 d_t 的建模。

在 $\ln P_t$ 分解过程中, ma_{t-1} 的移动平均项数选择,是以分解出的 d_t 序列为平稳序列为基本规则。通过数据试算,我们选定 2 期移动平均对 $\ln P_t$ 序列进行分解,根据平稳性检验结果,按照 2 期移动平均分解出的 d_t 为平稳序列。

本文选用沪深 300 股指期货的收盘价作为研究对象,样本的时间区间是 2013 年 7 月 22 日—2014 年 3 月 5 日,用作建模数据,将 2014 年 3 月 6 日—2014 年 3 月 21 日的数据作为模型验证数据。由于 d_t 序列存在 ARCH 效应,可以对 d_t 序列建立 ARCH 类模型。根据 AIC、SC 值最小原则,综合来看, GARCH (1,1) 为最优模型,具体建模结果为:

$$\begin{cases} d_t = -0.001 + 0.427d_{t-1} + \hat{\mu}_t \\ \quad \quad \quad (-120.580) \quad \quad \quad (310.127) \\ \hat{\sigma}_t^2 = -0.137\hat{\mu}_{t-1}^2 + 1.072\hat{\sigma}_{t-1}^2 + \hat{\epsilon}_t \\ \quad \quad \quad (-4.064) \quad \quad \quad (36.176) \end{cases} \quad (6)$$

依据 $\ln P_t$ 分解公式和上述 d_t 最优模型,还原出 $\ln P_t$ 模型:

$$\begin{cases} \ln P_t = ma_{t-1} + d_t \\ d_t = -0.001 + 0.427d_{t-1} + \hat{\mu}_t \\ \hat{\sigma}_t^2 = -0.137\hat{\mu}_{t-1}^2 + 1.072\hat{\sigma}_{t-1}^2 + \hat{\epsilon}_t \end{cases} \quad (7)$$

2.3 利用差分方法和序列分解方法模型的比较

预测效果比较可知,利用差分法建模预测的平均绝对百分误差 MAPE 为 69.24352,已高于 10,希尔不等系值为 0.596178,数值过大,而协变率为 0.352514,均方误差没有主要集中在协变率上。而利用序列分解法建模预测的 MAPE 为 0.104905,远远小于 10,预测精度很高,希尔不等系值为 0.000632,希尔不等系值越小越好,协变率为 0.964866,均方误差大多集中在协变率上,多种模型预测评价指标综合来看,使用序列分解方法建模预测精度更高。

3 股指期货套期保值研究

本文主要研究的是沪深 300 指数期货,假设 2014 年 2 月 19 日,某个投资机构具有市场价值为 1 亿元的华安上证 180EFT,当天的价格为 2.001 元/股,该投资机构预测未来股市将出现价格下跌,决定继续持有基金现货,并打算通过沪深 300 指数期货来对基金进行套期保值,对冲现货市场的系统性风险。根据上述的沪深 300 股指期货套期保值步骤,实施套期保值。

使用 OLS 方法计算最优套期保值比率,进行套期保值操作。根据 R_{hs} 和 R_{eft} 的回归分析结果,可知回归方程为:

$$R_{hs} = 8.47e^{-5} + 1.0548 \times R_{eft} \quad (8)$$

其中 $R^2 = 0.9565$,可知该回归方程拟合效果很好, β 值为 1.0548,即最优套保比率为 1.0548,计算股指期货合约数量为 153 手。

我们利用了 OLS 方法使用了现货的收益率对期货的收益率做回归,估计得到的系数就是我们所要的最优套保比率,这个方法十分简单,但是也有很多不足之处:

首先,我们没有考虑基差风险对于计算套期保值比率是否有影响,若现货的价格和期货的价格存在着协整关系,那么我们使用 OLS 方法就忽略了基差风险的影响。

其次,忽略了历史数据对于最优套保比率的影响,再者忽略了现货价格与期货价格两个序列是否存在条件异方差和自相关。

使用 ECM 方法计算最优套期保值比率,进行套期保值操作。根据协整分析表明:2014 年 1 月 6 日至 2014 年 2 月 18 日期间 $\ln EFT_t$ 和 $\ln HS_t$ 两序列间存在着协整关系,并确定误差修正模型:

$$\Delta \ln EFT_t = 1.06817ECM_{t-1} + 0.908211\Delta \ln HS_t \quad (9)$$

根据实证研究的 ECM 模型可知,最优套期保值比率 $\beta = 0.908$,计算得到股指期货合约数量为 131 手。

利用最小二乘法计算最优套期保值比率虽然简单,但存在了很多不足,既忽略了基差风险和历史数据的影响,也忽略了现货价格与期货价格两个序列是否存在条件异方差和自相关;利用误差修正模

型计算最优套期保值比率,考虑了现货价格和期货价格的历史信息,还考虑了滞后一期基差风险对于最优套期保值比率的影响,计算得到的最优套期保值比率更加准确,实际进行套期保值操作时规避风险更加准确。

4 股指期货套利研究

我们选择沪深 300 股指期货的两份不同到期日合约的差价来建模,利用差价建立的套利模型是当两份股指期货合约的差价波动高过均值一定范围的时候,选择做多差价;或者当两份股指期货合约的差价低过均值一定范围的时候,选择做空差价。但是随着时间的推移,根据价值规律,差价的波动朝向反方向发展,此时我们在进行反向的操作,通过这个过程获利。

4.1 选取数据

本文选取五分钟高频数据,能够提高套利模型的精确度,信息涵盖量更多。选取的建模数据为 IF1506 和 IF1509,时间区间为 2015 年 3 月 20 日 9:20 至 2015 年 3 月 27 日 15:15,选取的预测时间区间为 2015 年 3 月 28 日 9:20 至 2015 年 4 月 1 日 13:45。

4.2 套利模型

对两个不同到期日的沪深 300 股指期货即 IF1506 和 IF1509 的差价进行建模,用 ΔP 表示股指期货合约 IF1506 和股指期货合约 IF1509 的收盘价的差价,用公式表示: $\Delta P = P_{IF1509} - P_{IF1506}$ 。我们运用时间序列法建立模型,具体为

$$R_t = f(\ln \Delta P_t, \ln \Delta P_{t-1}, \dots) + \epsilon_t \quad (10)$$

根据 AIC 和 SC 最小准则、回归系数的显著性检验,经过反复的建模和检验,选择的模型如下:

$$R_t = 0.094 - 0.639 \ln \Delta P_{t-1} + 0.33 \ln \Delta P_{t-2} + 0.275 \ln \Delta P_{t-3} + \epsilon_t \quad (11)$$

根据异方差性和正态性检验,我们认为模型的残差项分布和正态分布没有显著差异。该序列服从均值为 0,方差为 0.063217^2 的正态分布,即 $\epsilon_t \sim N(0, 0.063217^2)$ 。

4.3 价格条件 VaR 的测算

根据我们上文建立的模型可知,IF1506 和 IF1509 差价的收益率估计值是:

$$\hat{\mu}_R = 0.094 - 0.639 \ln \Delta P_1 + 0.33 \ln \Delta P_{t-2} + 0.275 \ln \Delta P_{t-3} \quad (12)$$

从正态性检验结果可知,两合约差价的序列标准差估计值 $\hat{\sigma}_t$ 为:0.063217。

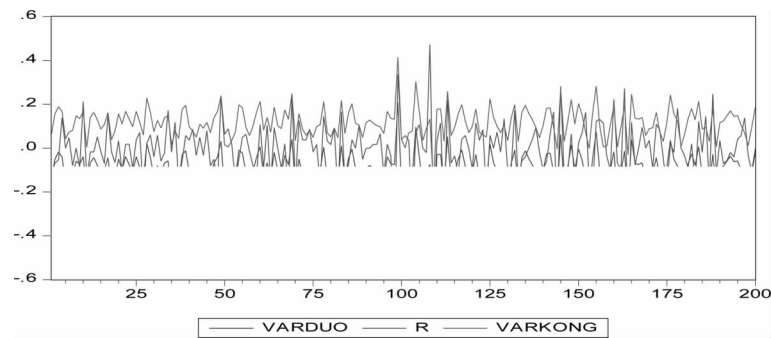
当 IF1509 合约和 IF1506 合约的差价高过均值一定范围的时候,选择做多差价,即买入 IF1509,并卖出等量的 IF1506,将置信度 α 确定为 0.05, VaR_t^{\otimes} 值为:

$$VaR_t^{\otimes} = \hat{\mu}_t - Z_{\alpha} \hat{\sigma}_t = \hat{\mu}_t - 1.644854 \times 0.063217 = \hat{\mu}_t - 0.10398274 \quad (13)$$

当 IF1509 合约和 IF1506 合约的差价低于均值一定范围的时候,选择做空差价,即卖出 IF1509,并买入等量的 IF1506,我们将 α 值确定为 0.05, VaR_t^{\otimes} 值为:

$$VaR_t^{\otimes} = \hat{\mu}_t + Z_{\alpha} \hat{\sigma}_t = \hat{\mu}_t + 1.644854 \times 0.063217 = \hat{\mu}_t + 0.10398274 \quad (14)$$

我们要预测 2015 年 3 月 28 日 9:20 至 2015 年 4 月 7 日 13:55 这五个工作日的 VaR 值,一共 200 组预测数据。我们利用上面的套利模型得出下面的预测值与真实值的折线图,它们的关系如图:

图1 VaR 与 R_t 比较

图中, R_t 是真实的收益率, 是风险测量的下限值, 是风险测量的上限值。从图中我们可以看到 VaR 值的预测区间, 3 月 30 日到 4 月 7 日这三个工作日内 200 个收益率的真实值落在预测区间内的概率十分大, 因此我们可以判定此模型建立的比较成功, 预测结果是相对有效的。

4.4 最佳套利时机的选取

根据正态分布检验 Shapiro-Wilk normality test, $W = 0.9913$, P 值 = 0.2779, 认为样本数据服从正态分布, 如拟合图 2:

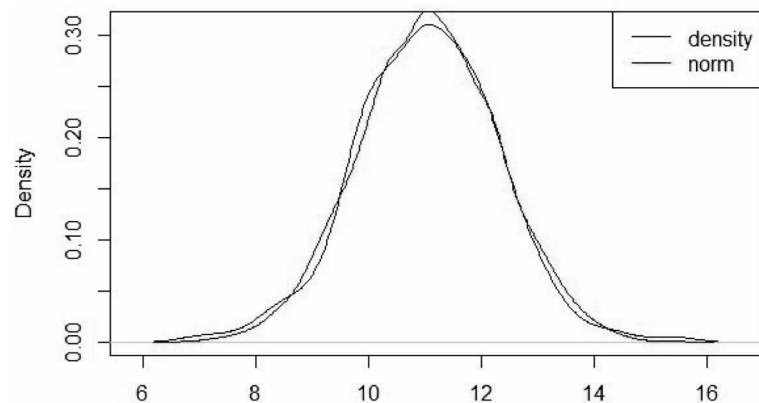


图2 拟合图

根据套利期望收益函数:

$$f(\alpha, Z_{\alpha/2}) = 600 \times \alpha \times Z_{\alpha/2} / 2 \times \sigma \quad (15)$$

利用 r 软件, 求得使套利期望收益函数最大的 α 值为 0.226, 如图 3:

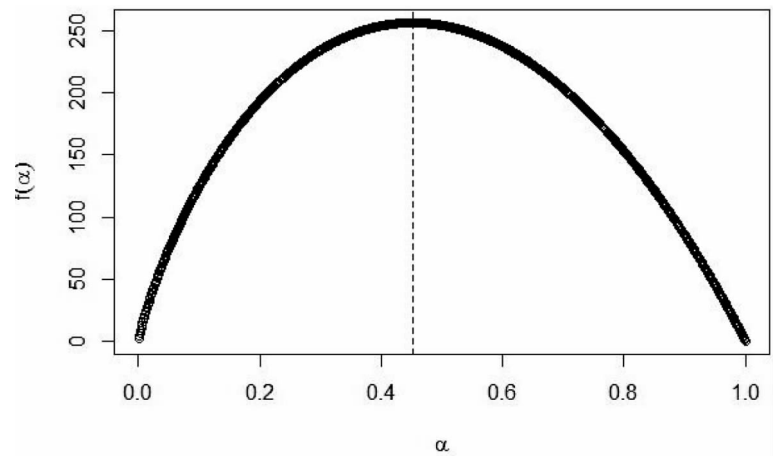


图 3 套利期望收益函数图

根据槐真(2013)¹⁶的最优套利点公式为:

$$Dn=\mu+Z_{\alpha/2}\sigma,Up=\mu-Z_{\alpha/2}\sigma \tag{4-7}$$

其中 $\mu=11.038, \sigma=1.258, \alpha=0.226$, 我们可以计算出做多点 Up 和沽空点 Dn , 其值分别为: $Up=10.09203, Dn=11.98397$, 最终套利区间图示为:

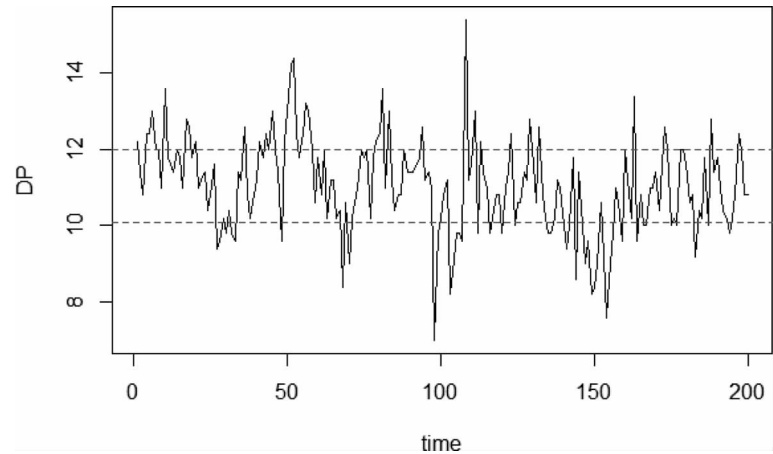


图 4 套利区间图

根据已知的两合约价差 ΔP , 做多点 Up , 沽空点 Dn , $VaR_t^{多}$ 值和 $VaR_t^{空}$ 值, 综合得到图 5:

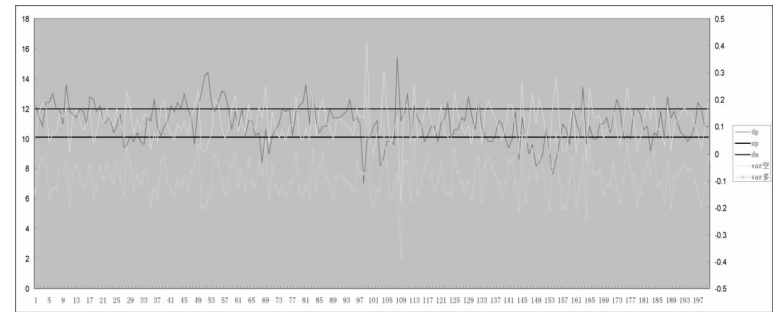


图 5 寻求最优套利点

我们从图 5 可知,两合约价差值 ΔP_t 最小时,最适合做多操作,随着 ΔP_t 越来越大,做多的风险值越大,做空风险值越小,越适合选择做空操作,当两合约价差 ΔP_t 达到最大值时,最适合选择做空操作。

参 考 文 献

- [1] William J. Baumol. An Expected Gain-Confidence Limit Criterion for Portfolio Selection[J]. Management Science, 1963, (10):174—182
- [2] Darrell Duffie, Jun Pan. An overview of Value at Risk[J]. The Journal of Derivatives, 1997, (3):7—49
- [3] Philippe Jorion. Value at Risk: The new benchmark for managing financial risk[M], The third edition. Oversea Publishing House, 2005
- [4] 王春峰. 金融市场风险管理[M]. 天津大学出版社, 2001
- [5] 郑文通. 金融风险管理的 VaR 方法及其应用[J]. 国际金融研究, 1997, (9):58—62
- [6] 肖春来, 宋然. VaR 理论及其应用研究[J]. 数理统计与管理, 2003, (2):6—10
- [7] 戴国强, 徐龙炳, 陆蓉. VaR 方法对我国金融风险管理的借鉴及应用[J]. 金融研究, 2000, (7):45—51
- [8] 肖春来, 柴文义, 扬威. 条件 VaR 理论的应用与研究[J]. 数理统计与管理增刊, 2003, (10):264—268
- [9] 肖春来, 柴文义, 刘喜波. 基于二维正态分布的条件 VaR 研究[J]. 数学的实践与认识, 2011, (06):171—172
- [10] 杨桂君, 肖春来, 常京宏. 基于二维 t 分布的 VaR 研究[J]. 数学实践与认识, 2014, (06):104—107
- [11] 姚芳. 收益率条件分布下的沪深 300 投资组合研究[D]. 北京:北方工业大学, 2012
- [12] 戴峰, 梁玲. 金融衍生商品市场中跨期套利交易的风险极小化模型与方法[J]. 系统工程理论方法应用, 1999, (4):7—14
- [13] 陈泽涛, 刘小红, 陈春芳. 基于 VaR-GARCH 模型度量股指期货风险—基于金融危机角度实证分析[J]. 商业经济, 2010, (22):23—25
- [14] 肖春来, 丁绍芳, 洪媛. 条件收益率下的 VaR 分析[J]. 北方工业大学学报, 2003, (3):88—94
- [15] 肖春来, 柴文义, 章月. 基于经验分布的条件 VaR 计算方法研究[J]. 数理统计与管理, 2005, (5):92—95
- [16] 槐真. 基于价格条件 VaR 的股指期货套利研究[D]. 北京:北方工业大学, 2013

Hedging and Arbitrage of Stock Index Futures Based on the Price Conditional Theory

Yang Guijun

(School of Economics and Management, North China Univ. of Tech., Beijing, 100144)

Abstract The CSI 300 stock index futures as the launch of the first branch of the stock index futures, the stock market has great significance for us. Since the transfer of risk itself stock index futures, price discovery and optimizing resource allocation features, we use these features for effective hedging and arbitrage operations. This article is on the Shanghai and Shenzhen 300 stock index futures hedging and arbitrage-depth study, first, using differential method and sequence of decomposi-

tion of the Shanghai and Shenzhen 300 stock index futures daily closing price for the number of modeling analysis, forecasting day closing price trend, and in turn evaluate the predictive effects of these two methods; second, the least squares method and the error correction model are two ways to solve the optimal hedge ratio, calculated according to the number of stock index futures contracts to solve the optimal hedging ratios, and hedging operations, and to assess the actual effect of hedging; and third, according to the price conditions for arbitrage model of VaR risk measure, and find the best arbitrage opportunity.

Key words stock index future; hedge; arbitrage; price conditional VaR

张伯苓“公能”教育思想研究

研究生 王 钰

导师 项 进

(北方工业大学马克思主义学院 北京 100144)

摘 要 中国现代教育家张伯苓的“公能”教育思想凝结于“允公允能,日新月异”的八字校训。“允公允能”,是张伯苓经过一系列教育实践,经历长期探索而确立的学校校训。他从中国国情出发,将新学办学方式与中国社会的现实及中国高等教育的实际情境相结合。“公能”教育思想主要包括“允公”教育思想和“允能”教育思想两方面。“允公”偏重公德教育,着重于对受教育者民族自觉和社会责任感的培养;“允能”偏重能力教育,着重于个人全面能力和现代能力的培养。

关键词 张伯苓;教育思想;允公;允能;“公能”教育

作者简介 王钰,女,1988年出生,硕士研究生,主要研究方向为思想政治教育。

E-mail:seasonyu1988@126.com

1 绪论

20世纪50年代以来,有关张伯苓教育思想的研究,未得到应有的重视,直至80年代,才引起了学术界广泛的关注,陆续出现了大量的研究专著和学术论文,可供研究的资料也越来越丰富、全面和真实。以王文俊为代表的致力于研究南开大学的发展历程和以梁吉生为代表的致力于研究张伯苓的教育思想的论著,都反映了关于张伯苓的研究日渐趋向系统化、全面化和客观化。但是,在这些成果之外,相关研究还有进一步的提升空间。首先,除梁吉生的专著对张伯苓的教育思想有系统的研究外,多数研究都偏重于论述张伯苓的生平,尤其是尚未形成有关“公能”思想比较系统的研究。第二、在学术论文的研究中,对张伯苓的思想的研究,论证不够深入和系统,也没有对张伯苓“公能”教育思想中的“允公”与“允能”之间的关系进行探讨。第三、相关材料大多集中在专门性的研究著作,少有张伯苓亲友、南开学生和西南联大学生的回忆性文章。本文尽可能搜集以往被忽视的一些零碎材料,丰厚对张伯苓“公能”教育思想的研究基础,以此从多方位全面深入地了解其思想。

2 张伯苓“公能”教育思想的形成与发展

张伯苓是一位功勋卓著的大教育家,是20世纪中国最杰出的教育家之一,其功可比蔡元培、梁启超、黄炎培、陶行知,他用尽毕生贡献于教育,形成了“公能”教育思想,他百折不挠,高瞻远瞩,创办的南开私立学校跻身清华、北大这些一流大学之列。张伯苓及其教育思想是社会改造和经济发展的强大助力,推动了时代的民主化进程。研究张伯苓的“公能”教育思想,需要从张伯苓“公能”教育思想的形成与发展着手,论证其教育思想的主要内容和实践途径,在此研究的基础上给予客观的评价,并从中获得启示。

2.1 张伯苓“公能”教育思想的形成

中国现代教育家张伯苓的“公能”教育思想凝结于“允公允能，日新月异”的八字校训。“允公允能”，是张伯苓经过一系列教育实践，经历长期探索而确立的学校校训，他从中国国情出发，将新学办学方式与中国社会的现实及中国高等教育的实际情境相结合。当时国家积贫积弱，国人愚昧自大，自私自公，如一盘散沙，缺乏团结力没有凝聚力，亡国灭种之灾无情降临在面前。作为一个爱国主义者，张伯苓根据时局状态和历史发展的要求，摸索实践着“允公允能”的教育理念，最终形成“允公”和“允能”的教育思想。以“允公”和“允能”作为其教育救国，实业兴国的实践标杆，旨在培养既有公德又有能力的人才，从教育的层面化除愚、弱、贫、散、私的社会顽疾。

“公能”教育思想是时代的凝结、智者的传承和教育智慧的精华，他的思想源头，主要是受到了儒家传统文化的熏陶、国外先进经验的感染、国内同仁实践的启示和天津本地同行的影响。张伯苓深受儒家传统文化的影响，其早期的学习经历多与儒家传统文化交臂。他十分关注中国传统文化，也真正理解传统文化的内涵，正确提出继承与发扬的关系问题。“公能”核心教育思想根源于中国传统文化，他还曾就儒家孝义与公民道德作出解释，诠释出公德不仅仅是个人品德的修养，而是社会公共道德的修行。张伯苓的“公能”教育是在亲眼目睹国外教学情景，亲身感受国外教育的先进中形成的，他的教育思想和办学实践都深受国外先进理念的感染，而其中的通识教育理念、现代教育方式和实用主义更是其思想的精华来源。在同一个特殊的时代下，梁启超、陶行知、蔡元培、晏阳初等国内同仁对张伯苓初期教育思想的形成有很大影响，因“碰撞”而发展丰富。张伯苓的“公能”教育思想与这些大教育家们出而合辙，他们的思想也在无形中互相获得给养。19世纪末20世纪初天津发生了一些社会变革，为仁人志士的教育事业营造了良好的社会氛围，张伯苓与这些志士交游切磋，不仅以自己的思想、识见影响于人，也从先辈和同侪身上汲取力量，激发教育灵感，激活内在动力。

2.2 张伯苓“公能”教育思想的发展历程

张伯苓“公能”教育思想经历了一个长期的探索完善的过程，历经了萌发期(1898—1918)、发展期(1919—1937)、成熟期(1938—1949)三个阶段。

1898—1918年是张伯苓摸索办学，为创办南开学校打下基础的第一阶段，是张伯苓“公能”教育理念萌发的时期，主要集中体现在早期南开中学全面型人才的培养模式上。南开中学不仅重视学生的智育开发，更强调学生的德育、体育和美育建设。张伯苓的早期德育教育思想从形成之初就相对成熟，虽然缺乏更深度的探讨，但究其广度，已经相对全面了。张伯苓“公能”教育理念经过了20年得以形成，到1919年他创办南开大学，又开启了新的教育篇章。

1919年至1937年张伯苓的“公能”教育思想得到发展。在“公能”教育思想的内容上，“公”的思想扩大化，实现本土化，“能”的思想也更加深化，“公”和“能”紧密联系在一起；在“公能”教育思想的实现方式上，逐渐从模仿办学走向独立办学的道路，从国情出发，从社会和人才的实际情况出发，为社会谋求发展，为当地发展作贡献。

1938—1949是中国历史的特殊时期，也是“公能”教育思想的成熟期，其成熟，表现在张伯苓“公能”教育思想的成果在西南联大三校合并时期得到共享和发扬。这个新的教育实体，坚持民主自由理念，强调兼容并包，注重德智体全面发展，培养了大批享誉世界的杰出人才，为战时的经济、科学和文明的发展输送了力量。西南联大时期，是张伯苓整个“公能”教育思想的成熟期和升华阶段。

3 张伯苓“公能”教育思想的主要内容

张伯苓倾尽一生投身教育事业，集萃而成的“公能”教育思想主要包括“允公”和“允能”两方面的内

容,“允公”是“公能”教育思想的核心,使受教育者爱护群体,有为公牺牲的精神,“允公”则是以“公”为基础,培养为公服务的能力。

3.1 “允公”教育思想

“允公”是张伯苓教育的基础,是其整套教育理念的出发点和落脚点。只有在教育中将“允公”放在首位,才能树立正确的价值观、人生观、教育观、学习观,只有在正确的“允公”教育下,才能进一步使受教育者“去愚去弱”,使国家“去弱去贫”,使人民的“能”有所用,有其落脚点。

“允公”包括“爱国”公德和“爱群”精神两方面。现代的中国需要的教育是要为社会谋进步,为公众谋幸福的。现代公民应具备强烈的民族自觉和责任感,要站在国家利益和民族文明的高度,打破孤立的本位民族视野,将时代精神和民族精神融合起来,为整个社会做出贡献。这种“大公”在几千年的教育思想中是有革命性的意义的。“允公”不仅要有“爱国”公德,还应具备“爱群”精神。“爱群”表现为群体意识。在张伯苓“允公”教育思想中,非常重视个人与群体之间的关系。这种爱群“公德”深合集体主义培养观念和现代人格的内涵,也是解决社会内部矛盾的钥匙。在张伯苓的整个教育理论和实践中,这种国家群体为先的公德教育都贯穿始终。

“允公”除了要求人们树立爱国爱群,服务国家的公德,还要求人们在现实中做到服务社会,使教育为社会谋进步,为地方谋发展。张伯苓在“允公”的教育思想中,独树一帜地突出了办学与天津本地发展之间的关系,致力于促进地方教育事业、地方实业的发展以及促进地方风貌改良。“教育为社会谋进步”是对受教育者的集体主义教育,也是对受教育者的理想教育。张伯苓从改良社会,反哺天津,移风易俗三个方面,诠释着他的“允公”教育思想,宣扬了与时俱进理念和创新精神,传播了集体主义教育思想,南开学子在受教过程中,也逐渐树立了科学的理想,为实现其社会价值,服务社会提供支撑。

3.2 “允能”教育思想

“允能”,主要是培养学生具备为国家为社会服务的能力,这种能力包括丰富的内涵,主要可以概括为以“德”、“智”、“才”、“技”、“体”为主要方面的全面能力和以科学知识和民治精神为主的具备实用价值的现代能力。

“允能”要求受教育者的全面能力的发展,培养人的全面能力,培养社会需要的专业型人才,在张伯苓所主张的全面能力的建设中,主要可以归纳为德能、智能、才能、技能和体能这几大方面。“现代能力”包括现代化的知识结构、民治精神和创新开辟的“现代”精神。科学知识和科学的知识结构是“现代精神”的基础和载体,没有科学知识,没有丰厚多层次的知识结构,“现代精神”仿佛“纸上谈兵”,是没有实际意义的;“现代精神”又是获取科学知识和民治精神的内因,如果说科学的知识体系是一种外化力量,那么“现代精神”就是内核,而服务于社会的实用价值就是“现代能力”的归依和最终落脚点。只有真“能”,才能“去愚去弱”,有为公服务社会的现实能力。

3.3 “允公”与“允能”的关系

“允公”是培养人才的起点和先导,是教育的基石。作为人才培养和发展的核心,决定了人才发展的前途方向和基本趋势,它是立国育才的精髓和灵魂,解决了为谁培养的问题。“张伯苓开创教育的源动力是爱国,教育是从爱国开始的,教育理念和实践的发展也是根据爱国的升华而逐渐变化的。因此,爱国救国、服务社会是“允公”的主要精华,是其人才培养的首要目标,“能”应服务于“公”,服务于当地实业。天津著名的塘沽久大精盐厂、永利制碱公司等大型化工企业,是“公能”教育的受益者。南开的商科、电气工程系、化学工程系等理化系不断增设与社会和市场相关的科系,与时俱进地调整,也培养了适合中国环境和发展之路的人才。

“允能”是人才发展的具体实践途径,是人才发展的要求,也是人才培养的重点。“允能”是“允公”

的真实要求和实现方式,为“允公”提供了具体实践模式和条件。“能”,主要是在于培养学生的为国家为社会服务的能力,这种能力既包括现代化的理论能力,又包括实际工作的能力,在德、智、才、技、体各方面全面平衡,尤其强调培养以现代科学的知识结构、民治精神、“现代精神”为主要内容的现代能力。通过这种为“公”服务的社会能力,使民众强大智慧,使新一代青年有才干面对社会和世界,使中国人才走向现代化。

南开“允公允能”的教育思想,有着深厚的理论源泉和根基,是在深切发掘中华建设之需要和把握南开全面人才培养的基础上提出的新式教育观,它与时俱进,在实践和时代中树立教育理念,摸索人才发展和社会进步的规模,凝练出新的教育方略。

4 张伯苓“公能”教育思想的实践途径

张伯苓的教育思想不仅仅在理论上有所建树,还形成了现实性的指导,形成了相对成体系的“公能”教育思想的实践途径,包括专业科系的特色设置、教师队伍的建设、办学经费的筹措和使用、管理体系的形成等全方面多层次的实践途径。

4.1 讲求实用的专业设置

南开大学在实践教育思想上,贯彻院系设置的“实用性”功能,注重科系建设的实用性以及研究机构的实用性。无论是商科还是理工科,张伯苓在科目的设置上都致力于实用,开设应用性课程、充实教学设备、加强理论与社会实际的联系、为学生创造进行社会实践的条件。他将实习课程设置成教学的一个重要环节,让学生融入到社会学习,为国家解决实际问题。张伯苓认为,中国教育的弊端在于学不能致用,学生最大的问题在于不了解社会。据此,张伯苓除了设置实用性的专业科目以外,还专门设立研究社会实际问题的东北研究会、经济研究所和应用化学研究所这样的研究机构。

4.2 学行并重的教师队伍

张伯苓以建设一支学行并重的教师队伍为己任,重视教师人才的选聘,重视教师队伍的建设,为南开培养未来的教师储备力量,为教师教书育人营造良好的教学环境。

张伯苓选聘教师既重视专业技能,又强调教师德才兼备,反对只注重“个人天才”。选聘中,必须注重教师学识和思想两方面,学行并茂才能入选,只有如此,才能对受教育者产生真正的影响,实现真正的“公能”教育。这种“有德有能”的要求反映了南开师资队伍建设的特色。他还认为,教师仅靠一人之力是不能成就好学校的,只有汇聚无数优质的教师,组成一支优秀的教师队伍,才能让所有学生都被塑造造成最杰出的青年。南开以学科带头人为核心领导,形成教授团队,既反映了现代社会的要求,顺应了现代大学的发展趋势,又遵循了“大公”的实际要求,革除私、散之弊。南开教师队伍的建设,重视教师的行为道德修养。张伯苓提出对于“行”的教育,“行”不仅限于“做”,也指向“行为道德”,“学”即使十全十美,“行”这一方面若不能秉珍自重,仍无法真正为人师表。知识科学应当是教师教育的教授任务,而教授学生做人的德行,行做人之道,帮助学生提高思想素质,增强公德心,培养集体兴趣和团结合作观念,才是矫正时弊的关键。如此教师建设正适合南开的发展,促进了中国教育事业的发展。

4.3 多渠道筹措办学经费

经费是一个大学赖以生存和延续的基础,南开作为一所私立大学,其经费更是办学发展的最大桎梏。张伯苓通过拓宽渠道寻求经费,多种方式筹措经费,恪守原则善用经费,他的“公能”思想奠定了相对稳定的施展平台,也成就了他的办学梦想。

南开的经费主要来源于学生缴费、校产及基金利息、政府补助、社会捐助,社会捐助和后期的政府

补助是其经费来源的主要构成。张伯苓多方拓宽经费来源,利用校董会、专项募款活动、捐启、学术项目、开展社会教育、有偿科技服务和科技成果商业化来增加经费。这些多渠道的经费来源,一方面说明了社会对“公能”教育理念的认同,从侧面反映了南开在社会中的影响力。另一方面,由于“南开是社会的产儿”,正因此,南开对社会做出了回报,培养人才,开发科研成果,以此“服务社会,服务当地”。南开的经费多是张伯苓“求”来的,他跑遍四处,各方求人说项,甚至在向官僚军阀募款时,还遭到学生的强烈不满,他所承受的压力和白眼是一般大学校长很难经历的。这样难得的“经费”,张伯苓始终做到公开、透明、无私,让人动容,他通过自身的德行,影响着周围的人,甚至是整个地区。这种“允公”的教育完全融化在他的行为实践中,比直接教学带来了更为深远的影响。张伯苓在经费问题上诠释着他的办学理念,无论从经费的来源、方式和影响,甚至是从张伯苓的德行来看,无不在诠释他“允公”“允能”的教育理念。

4.4 特色鲜明的管理制度

在管理上,张伯苓重视从民主德治化管理和严格法治化管理两方面进行建设。

“力合则效果大,力散则效果微”。南开要办成一种“家庭学校”,不是为管理而管理,而应该和谐如一个大集体,宽严相济,顺应自然发展,发挥管理的杠杆作用。民主与德治的联系源于张伯苓对传统道德真意的延伸阐述。他将传统道德上升到公德层面,将西方的民主观念与“德”相结合,诠释出新的道德;站在世界的角度,蕴含民主、平等意义的“德”。张伯苓针对20年代中国教育在管理上的弊端,将这种德与民主的关系通过学校管理的实践加以演绎,实行师生共治,民主管理,同负责任,普遍地增强人们心中的“公”的观念,团结的观念。南开的管理不仅处处体现人文思想,民主自治管理特色,也树立了法治化管理学校的典范。学校将主观的意识落实到客观的统一化公开的管理规定上,建立健全各种规章制度。张伯苓的管理意识,营造了良好的学校教育氛围,为南开赢得了近20年的相对独立的发展,奠定了南开的历史基础和历史价值。

4.5 国内领先的体育教育

张伯苓对体育的见解高瞻远瞩,发人深省,直至今日,仍然有很强的现实意义。他从国家立场出发,提出“强国必先强种,强种必先强身”;从教育与体育的关系出发,认为体育是教育的必要部分;从培养人才的角度出发,主张培养德智体的全面人才;从教育者的角度出发,认为“不识体育者,不应该做学校校长”;从体育的内部关系出发,要求体育技能和体育道德相长。张伯苓将其体育教育思想付诸到整个教育实践上。

4.6 艺术教育与人格塑造

南开高度重视积极推广新剧教育,强化艺术教育以培养全面人才,重视通过艺术教育以实现人格塑造。南开的艺术教育和美育教育思想是“公能”教育思想的具体要求和实践途径。南开学生可以充分发挥个性特长,满足自身个性发展和主动发展的需要,逐渐形成社会责任感、创新意识、实践能力、高尚的情操和高雅的文化素养,符合时代精神的需求。

5 张伯苓“公能”教育思想的评价与启示

“公能”教育思想是张伯苓教育思想的核心,也是张伯苓办学的方向标。研究张伯苓“公能”教育思想,论证其在办学理念、治学风格、大学定位、发展方针、教育宗旨和人才培养上的理论和实践价值,全面深入地研究张伯苓“公能”教育思想,剖析张伯苓“公能”教育思想的历史影响,有利于充分体会其思想的现实价值。

5.1 张伯苓“公能”教育思想的评价

张伯苓经过 41 年的实践,历经坎坷,结合西方先进教育经验与中国本土教育情况,形成了“公能”教育思想,不仅树立了对民办教育的指导,也形成了当时教育界的理论高峰。

张伯苓“公能”教育思想经过不懈的教育实践,取得了巨大的成就,形成了成熟、系统的理论体系和应用经验,传播了“允公”的核心思想,建立了系统的“公能”教育实践和理论体系,促进了天津本地的发展,成为抗战的重要力量,培养了大批德才兼备、遐迩闻名的人才,探索出了适合中国国情和社会发展的教学模式。更值得重视的是,独特的体育和话剧成就写下了光辉的历史,对当代乃至后代都做出了突出的贡献;张伯苓的体育教育思想无论在硬件建设,在体育精神和风气,在体育人才和体育事业上,都取得了巨大的成就,他还将体育教育和成果推向世界,将中国的风貌推向世界,培养了面向世界的全面型的现代体育人才,推动形成了一种团结观念、国家意识和奉献精神,在全国产生了广泛的影响,是真正的奥运第一人,真正的体育教育家。张伯苓率先将话剧引入学堂,并且利用话剧对人们进行现实教育,用隐形的手段培养人们的欣赏水准,提升受教育者的人文素质和修养,通过艺术的熏陶进行人格感化,培养受教育者的全面素质。话剧作为精神诱导的良剂,被南开成功地用来改良社会,对立事处世,收获良多。

5.2 张伯苓“公能”教育思想的启示

“公能”教育思想对当代新时期的高等教育仍然有着启示作用,可以说有功于历史,有益于当代。通过分析和研究张伯苓“公能”教育思想的历史影响力以及现代价值和地位,为新时代高等教育的转型、普通高校的定位和人才培养的模式提供理论依据,为普通高校转变塑造尖端型人才,转而打造服务于社会的中间型人才提高实际指导,为培养中国现代公民的内在素质和人格提供理论支撑。正确认识张伯苓的“公能”教育思想并且给予客观的评价,可以使我们从中获得启示,推动我国当代教育事业的发展,对新时代下实施素质教育、增强学生的社会责任感、培养学生的创新精神和实践精神,增强学生的就业实力,树立法治民主的管理体系有着普遍的指导意义。研究张伯苓的“公能”教育思想,可以丰富高等教育的理论体系,为新时代高等教育的转型和人才培养提供理论来源。

参 考 文 献

- [1] 梁吉生. 允公允能 日新月异——南开大学校长张伯苓[M]. 济南:山东教育出版社,2006:138
- [2] 王文俊,梁吉生. 张伯苓教育言论选集[M]. 天津:南开大学出版社,1984. 247
- [3] 彭援军. 张伯苓:中国奥运的先驱者[J]. 北京档案,2008(2):13—16
- [4] 邢纯贵,江沛. 爱国教育家张伯苓的体育思想及其实践[J]. 南开学报,1996(5):75
- [5] 徐大军. 南开新剧运动的社会教育意识[J]. 艺术百家,2005(4):42—45
- [6] 詹根连. 张伯苓教育管理思想和实践探微[J]. 高等师范教育研究,1992(2):57

“Public Interests and Practicality Capability” : the Research of Po-ling Chang’s Educational Thought

Wang Yu

(School of Marxism, North China Univ. of Tech. , Beijing, 100144)

Abstract Chinese educator Po-ling Chang’s “public interests and practicality capability” educational thought originated from school motto: “Dedication to public interests, acquisition of all-round capability, and aspiration for progress with each day”. It took shape through long-term exploration and practice. He combined the western advanced educational experience with local educational situation together. “public interests and practicality capability” educational thought contains two parts, “public interests” educational thought emphasizes on public interests and morality education, which focuses on cultivating national consciousness and social responsibility, while “practicality capability” lays on quality-oriented education, which put emphasis on training individual comprehensive ability and modern ability.

Key words po-ling chang; educational thought; public interests; practicality capability; education of “public interests and practicality capability”