

文章编号: 1007-5399 (2014) 05-0011-03

# 电子商务小包名址自动采集系统的应用研究

黎达云, 孙举恒

(邮政科学研究规划院, 北京 100096)

**摘要:** 文章分析了建设电子商务小包名址自动采集系统的必要性, 阐述了电子商务小包名址自动采集系统的组成、工作流程及关键技术实现, 进而总结了该系统的应用效果, 提出了系统应用的前景展望。

**关键词:** 智能识别技术; 电子商务小包; 名址; 数据; 技术

**中图分类号:** F61

**文献标识码:** A

电子商务小包业务是中国邮政专门针对国内轻小件寄递市场推出的全新产品, 重点关注电子商务行业的各类寄递需求, 向协议客户提供批量交寄、预约投递、上门签收、投递过程短信通知等个性化服务。该业务自推行以来, 业务量迅速提升, 随之产生的名址信息录入工作给邮件投递时限和员工劳动强度带来极大挑战。自动化智能识别技术在降低劳动强度、缩短投递时限方面起到了重要作用。

## 1 电子商务小包名址自动采集系统的研究背景

电子商务小包的全网信息共享和网络化分拣的实现, 对收寄环节提出了更高的信息质量要求。按照生产作业要求, 在邮件封发前要先完成收件人邮编及名址信息录入, 这也是最耗时间的一个环节。目前 60% 左右的邮件通过淘宝等第三方数据接口或其他形式直接获取收寄名址信息, 其余 40% 左右的邮件商户不愿或无法提供电子收寄名址信息。以一个日均 2 000 件电子商务小包的处理局计算, 约有 800 件的收寄信息需要手工补录, 并且要求在小包到达处理现场后的短暂时间内完成, 否则难以保证及时赶上转运车次, 给营业部门带来了巨大的工作量。

如何利用自动化智能识别技术解决上述问题, 实现大宗电子商务小包名址和条码信息的高速、批量自动采集, 大幅降低员工信息补录劳动强度, 提高工作效率, 加快邮件投递时限, 降低人工成本, 已成为邮政部门大力发展电子商务小包业务急需解决的难题。

## 2 电子商务小包名址自动采集系统的需求分析

为做好电子商务小包业务发展的技术支撑工作, 邮政科学研究规划院针对电子商务小包名址和条码信息的高速、批量自动采集需求, 提出开发一种小型自动化地址采集系统, 综合利用图像扫描、光学字符及光学条形码识别技术, 实现电子商务小包面单的自动分离和图像采集, 并识别面单上的收件人名址和条码信息, 将信息导出供其他相关系统使用。

该系统应具备以下功能: 第一, 能批量供件, 自动实现面单分离、图像采集和信息识别。第二, 设备应小型化, 操作简便, 并具有较高的处理速度(速度应在 3 000 件/小时

以上)。第三, 能自动区分收件人的邮编、地址、姓名、电话以及邮件条码等信息, 具有较高的识别率。第四, 具有信息剔除功能, 能对无法识别的图像进行剔除。第五, 系统应具备审核补录功能, 能对采集的名址信息进行查询管理及导出, 并能对地址采集质量进行统计分析。第六, 系统适应性强, 能适应各电子商务根据标准格式打印生成的面单。

## 3 电子商务小包名址自动采集系统的组成及工作流程

电子商务小包名址自动采集系统由扫描设备和软件系统组成。扫描设备采用市场上的成熟产品, 可以实现 A4 幅面纸张的连续供件、自动分离、单/双面扫描等功能。

系统主要工作流程包括面单分离、数字图像扫描、图像处理及版面分析、光学字符及光学条形码识别、信息剔除、数据后处理、地址智能纠错、地址数据存储等, 如图 1 所示。

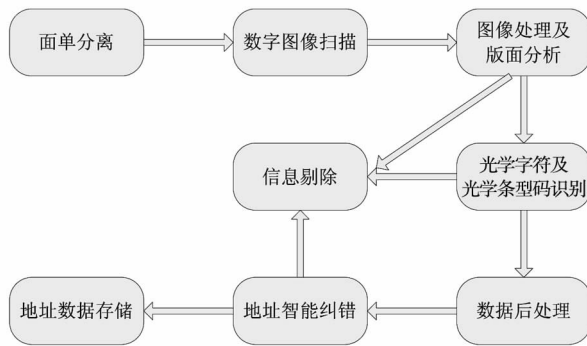


图 1 电子商务小包名址自动采集系统工作流程

面单分离和图像采集由扫描设备在软件系统控制下完成。面单分离是指将电子商务小包面单放到扫描设备的供件槽后, 在软件系统驱动下实现面单的自动分离。图像采集是指面单通过单式分离进入扫描设备的扫描通道后, 在指定存储路径生成规定格式的原始图像文件, 以供后续处理使用。

图像处理及版面分析、光学字符及光学条形码识别、数据后处理、地址智能纠错、地址数据存储等处理流程由软件系统完成。系统采用扫描设备完成数字图像扫描后, 通过先

进的图像处理及版面分析技术实现图像中条码区域和有效地址区域的自动判定,利用光学字符及光学条码技术完成条码及地址信息的识别,并对不能识别的图像进行信息剔除,然后根据电子商务小包面单上各地址信息域的分布位置规律,通过智能化数据后处理算法实现各地址信息域的数据拆分和拼装。考虑到光学字符识别结果会存在差错,系统进一步利用全国各县及县以上行政区划标准地名对收件人地址进行智能纠错,并对识别结果进行人工审核确认。审核通过的地址图像可以按照指定格式存储,并将地址数据导出到其他业务系统所需的文件格式。

#### 4 电子商务小包名址自动采集系统的关键技术实现

电子商务小包名址自动采集系统的开发涉及自动化控制、图像扫描、图像处理与版面分析、光学字符识别、光学条码识别、数据后处理、地址智能纠错等多项人工智能技术,具有较高的技术难度。目前,图像扫描、光学字符及光学条码识别技术已经相当成熟,印刷字体的识别率也达到了较高水平,但如何与业务单式相结合,实现较高的识别率,国内尚无统一的实现模式。尤其是各家电子商务的面单打印软件形式多样,打印格式不统一,给系统的信息识别带来了困难。若系统不能达到一定的准确度,地址的整体识别效果将大打折扣。为了达到最佳的识别效果,邮政科学研究规划院对面单打印格式规范化、图像预处理、全版面分析、数据后处理、地址智能纠错等关键技术进行了有益的创新尝试。

##### 4.1 面单打印格式规范化

尽管邮政部门已经在面单中规定了各类信息的打印位置,由于各家电子商务的包裹面单打印软件五花八门,造成不同电子商务的面单打印格式千差万别,各信息项之间的间隔距离也不尽相同。如何从面单图像中快速确定有效地址区域以供光学字符及光学条码识别使用,可以有多种不同的应用实现手段。最简单、直接的办法就是为每个电子商务的不同类型打印单式设置不同的面单区域模板,针对条码、邮编、电话、单位、地址等不同区域建立不同图像。由于邮政面向众多电子商务,需要为不同电子商务建立不同的版面分布图像。当需要建立不同模板的电子商务数量较大时,地址模板的维护管理将变得异常复杂,因此这种模式并不适用于电子商务数量较多的应用场合。从电子商务小包面单看,条码,收件人的邮编、电话、单位、收件人地址,寄件人的邮编、电话、单位、寄件人地址都有规范可循,它们有相对固定的打印位置和格式,因此可以据此建立一个规范化的电子商务小包面单通用打印格式,然后在各电子商务企业中推广使用,这样软件系统就能轻松地根据各信息项之间的位置特征,对各区域进行智能区分,这将大大缩减面单模板的配置数量。

##### 4.2 图像预处理

由于电子商务小包面单基本为针式打印机打印,打印字体易出现断线、边界模糊等现象,再加上复写打印还易产生

干扰污点,因此为确保数字图像具有较高的可识读性,必须对原始图像进行相应的预处理,从而去除图像位深度、背景灰度等方面的差异,实现图像自动纠偏、噪点去除,生成轮廓清晰、背景分明、无干扰的二值化图像。

##### 4.3 全版面分析

在图像预处理基础上,需采用多维扫描方式对地址图像中的条码区域进行分析,提高图像倾斜和条码污损情况下的条码识别率。根据电子商务小包地址“特征点分析”算法对图像中的多个地址区域进行智能分析,在识别图像中的条码区域后,进一步获取地址图像中的全部有效地址区域。

##### 4.4 数据后处理

在完成图像预处理和全版面分析后,系统利用光学字符、光学条码技术对标注的全部有效地址区域进行识别,根据电子商务小包面单上的邮件条码,收件人邮编、电话、姓名、单位、收件地址,寄件人邮编、电话、姓名、单位、收件地址,大客户条码等信息项的标准打印位置,各信息项的字符组成特征,各信息区域的行高和间隔等特性,利用智能化数据后处理算法将光学字符/光学条码识别结果逐个拆分到相应的地址信息项中。例如:邮件条码总是出现在图像的左上方,收件人邮编一般在邮件条码区域下面的第一行有效地址区域,邮编为6位数字字符,收件人电话、寄件人电话由不超过11位数字的字符组成,各信息域之间有明显的行间距,收件人地址最多不能超过3行等,诸如此类的特征信息均可作为智能化数据后处理的判断依据。

在光学字符/光学条码识别结果信息拆分的基础上,系统还根据错误驱动、汉字特征模型和历史统计学习等算法去除干扰字符,并对识别错误的阿拉伯数字和英文字母进行智能修正。

##### 4.5 地址智能纠错

使用针式打印机打印电子商务小包面单,易造成少数汉字被误识为其他字的现象,如:“州、川、新、乌”等,对于收件人地址而言,如果识别出来的省市县地名中存在错别字,将严重影响到该地址对应邮编的匹配及其后续应用。因此,邮政科学研究规划院开发了基于标准地名关键词的地址相似度智能纠错算法,较好地纠正了收件人地址光学字符识别结果中的地名差错,从而保证名址自动采集系统得到的县及县以上行政区划地名的准确率达到99.5%以上。

经过上述关键技术创新,较好地解决了电子商务小包地址识别过程中遇到的各种技术难题,经使用打印规范的电子商务小包面单实际测试,系统的条码识读率达到99.9%以上,多数不能识读的条码是因为条码被错位打印字符遮挡;邮编和电话的识别率达到99%以上,收件人地址中的县市地名识别率达到99%以上,整机处理速度为3500件/小时,取得了较好的识别效果。

#### 5 电子商务小包名址自动采集系统的应用效果

为使该系统能够真正满足邮政实际生产的需要,项目组在研发过程中广泛征求了中国邮政集团公司业务主管部门和

# 小议邮政便民站建设的优化策略

各地邮政部门为统筹城乡邮政均等化服务，拓展和延伸邮政渠道建设，完善和提高邮政公共平台打造，在各级政府和社会各行各业的大力支持下，把加快和加强邮政城乡便民站建设，作为政府服务民生的政绩工程及邮政企业的便民工程来抓，搭建城乡邮政服务“半小时”或“500米”便利圈。但是，各地邮政企业热衷邮政便民站建设的同时，在便民站的使用管理、服务质量、效能效果方面都存在滞后性，如使用率不高，服务功能单一，常态化管理缺位等，影响了邮政便民站服务社会形象和效能的提升。

要想更好地发挥邮政便民站点多面广、服务城乡的优势，必须坚持为政、助商、便民的服务理念，自觉遵循科学发展观为指引，全面落实邮政便民站便利、快捷、综合、高效的服务要求。从机制体制入手，服务需求着眼，转变思路考量，打造服务形象，以实现由量到质的转变，提升邮政便民服务站的服务价值和品牌价值。

## 1 规划好邮政便民站的布局选址

随着城乡建设步伐的加快，特别是各地在加快推进城市化建设的新形势下，做好城乡邮政便民站的选址布局尤其重要。各地邮政企业应立足当前、谋划长远，用科学统筹的眼光规划本辖区、本区域内邮政便民站的选址布局。

## 2 叠加好邮政便民站的业务项目

邮政便民站的建立是为了更好地为民众提供就近的邮政服务，因此邮政便民服务站在满足民众一般性的订报、寄信等需求外，还要进一步进行业务资源的整合，扩充业务服务范围、叠加业务经营范畴。

## 3 管控好邮政便民站的运营质量

在社会和企业的大力推动下，邮政便民站建立得到了

快速发展，成为城乡便民服务的一道靓丽风景线。为做到有序经营、有效管理，促进邮政便民站的稳健科学发展。抓好顶层设计，规划运营，结合实际制订科学规范的邮政便民站的经营、服务、管理公约，努力提高邮政便民站的运营质量，同时强化便民站的运营管控，进行定期和不定期的检查督导，发现问题及时查处，不断强化合规意识，提高服务水准。

## 4 培训好邮政便民站的加盟商户

为努力提高加盟商户整体水平，塑造良好的便民服务形象，必须进一步加大加盟商户的培训力度，通过集中培训、送教上门等形式，系统培训加盟商户业务发展、服务管理等方面的知识，全面夯实和提高加盟商户的能力。

## 5 宣传好邮政便民站的便利为民

邮政便民站的开办时间、社会影响还处在初期阶段，需要进一步扩大影响力和知名度。要在邮政便民站的标牌标识、业务办理、服务规范、品牌打造等方面加予提升和完善，让民众能够通过宣传了解邮政便民站的便利为民。可整合和应用社会电视、网络、广播等公共传媒，也可利用邮政自身的直邮广告、邮政彩信、数据商函等载体进行形式多样、家喻户晓的宣传推介。

## 6 理顺好邮政便民站的合作协调

由于邮政便民站的建立是在整合社会资源基础上完成的，因此理顺和协调好各方的合作关系，对提高邮政便民站的品牌打造意义深远。通过联席会议机制、签约合作协议等形式，在法律框架内，互惠互利，共同推进邮政便民站的又好又快发展。

(广东省梅州市邮政局 陈国红)

相关省市使用单位的意见，这些宝贵的改进建议对系统的功能提升具有较大促进作用。

在系统研制完成后，为了检验系统的实际运行效果，分别在保定、北京、广州等地的电子商务小包集中收寄部门对该系统进行了现场试用。在对相关电子商务的面单打印进行规范化调整后，电子商务小包名址采集数量和质量有了明显提高，并实现了较高的条码和地址识别率。现场工作人员普遍认为该系统方便实用，名址采集速度快，识别准确率高，能够较好地解决大宗电子商务小包名址采集中存在的问题及实际需求，大大降低了劳动强度，提高了工作效率，达到了提升邮政自动化水平、降本增效的效果。

## 6 电子商务小包名址自动采集系统的应用前景展望

电子商务小包名址自动采集系统的成功研制，为电子商务小包名址和条码信息的高速、批量自动采集提供了重要手

段。目前仍有相当比例的商户不愿或无法提供电子收寄名址信息，可以将该系统应用到营业网点、大客户中心等电子商务小包收寄机构，实现大宗电子商务小包地址和条码的高速、批量自动采集，这将是电子商务小包全名址采集的重要补充，对于大幅降低邮政员工劳动强度，加快邮件投递时限，具有重要意义。

## 参 考 文 献

1 龚永恒. 基于图像特征提取的手写体数字识别. 现代邮政, 2011, 3

作者简介：黎达云（1974~），男，广西陆川人，硕士，高级工程师，主要从事邮政信息化、标准化及流程优化和应用软件开发研究；孙举恒（1971~），男，河北冀州人，高级工程师，主要从事邮政信息化、标准化及流程优化和应用软件开发研究。