

降水自记纸彩色扫描数字化处理系统*

王伯民¹⁾ 吕勇平²⁾ 张强¹⁾

¹⁾ (国家气象中心,北京 100081)

²⁾ (广东省气象局,广州 510080)

摘 要

降水自记纸彩色扫描数字化处理系统是实现由降水自记纸彩色图像转变成每分钟降水量的计算机处理系统。该文介绍了降水自记纸彩色扫描数字化处理系统的设计思想、系统结构和系统功能(自记纸预处理、扫描、扫描检查、降水曲线提取、降水强度数据转换),并从降水自记纸扫描图形的数字化精度分析和扫描分辨率、降水曲线自动跟踪提取(降水曲线彩色表征区分、降水曲线迹线跟踪提取、图像旋转技术、图像二值化和滚动条改变阈值、图像同步显示技术、鼠标定位键盘方向键微量移动修正技术)、图形数字化数据处理技术(尾数累计进位及其误差、二项式滑动平均法)、同步定位显示技术等系统设计中的代表性技术进行了介绍。降水自记纸彩色扫描数字化处理系统已在全国各省投入业务应用。

关键词: 彩色扫描 图形数字化系统 降水自记纸

引 言

降水自记纸是客观记录降水情况的十分宝贵的气象实测资料之一。降水自记纸的降水曲线,客观记录了降水随时间变化的过程,反映了各种时段的降水强度,是其他降水观测资料所不可替代的。降水随时间变化的过程和降水强度,对于研究我国降水特征,尤其是暴雨特征具有重要意义,并且对农业、城市建设、电讯、交通、水利等国民经济部门,对防灾减灾策略的制定具有重要作用。由于降水自记纸的图形表现方式和纸张载体的限制,其信息无法得到广泛和深入的应用。因此,降水自记纸的数字化处理,是改变目前状态、使其信息的内在含义得到充分利用、价值得以充分体现的唯一途径。

中国气象局所属气象台站积累的降水自记纸,由于年代长,保管条件差,已经出现不同程度的纸张变质、字迹变淡(模糊)等现象,急需采取措施抢救。降水自记纸的数字化处理,是实现抢救目标的根本措施。

有关气象图形图像处理技术的研究和应用相当广泛,例如卫星和雷达图形图像的处理;相关的研究论文也很多,并形成了一些系统的著作和教材,例如国防科研试验工程技术系列教材之一的《气象图形图像处理与应用》¹⁾。但是,对于像降水自记纸这样从图像中剥离曲线信息的数字化技术,作者还没有发现国内外有关的技术文献和报道。

随着信息技术的不断发展和完善,将传统纸质的大量资料转变为数字资料,进行方便的存储、利用和传输已变为现实。彩色扫描和计算机图形处理技术的发展为降水自记纸

* 2004-01-18 收到,2004-07-16 收到修改稿。

图形的处理提供了一种有效的手段和方法。

实施珍贵气象资料拯救工程,是中国气象局《全国气象事业发展第十个五年计划》发展与建设的内容之一。为落实“十五”计划,预测减灾司安排小型建设项目,在全国组织开展国家基本(准)站降水自记纸的数字化处理工作并委托国家气象中心和广东省气象局研制降水自记纸数字化处理系统。

降水自记纸数字化处理系统的设计开发工作从 2001 年 4 月份开始,至 2002 年 11 月初步完成。首先在广东省、黑龙江省、山西省、宁夏自治区、四川省、江西省气象局进行试验,并于 2002 年 12 月开始在全国逐步推广应用。经过 2003 年一年的试用对“系统”作了进一步改进,降水自记纸数字化处理系统已经比较完善。

1 降水自记纸数字化处理系统设计思想和结构

1.1 降水自记纸数字化处理系统设计的基本思路和目标

采用先进的图形扫描和数据处理技术,实现对降水自记信息的完整、真实、安全的保存;提取降水曲线数据,形成全国地面基准、基本站降水强度数据标准文件;为全国一般站的降水自记和其他图形记录档案的数字化,奠定基础、储备技术。

降水自记纸数字化处理系统目标为:业务流程科学合理;自记纸数字化处理计算机系统操作简单,自动化程度高;形成清晰、完整的降水自记纸彩色扫描图像文件;形成高精度的降水分钟和小时强度数据标准文件。

1.2 降水自记纸数字化处理系统结构

- (1) 设备配置 由 1~3 台微机和 1 台 A3 幅面彩色扫描仪组成。扫描仪型号:EPSON / GT/10000+;微机配置:处理器(奔腾 III 933 MHz),内存(128 M),硬盘(40 G),光驱,显示器,刻录机。
- (2) 软件环境 操作系统:Windows 98/2000;EPSON 扫描仪软件,Page Manager for EPSON 软件。
- (3) 系统结构
 - ① 降水自记纸数字化处理系统由 4 部分组成(图 1):

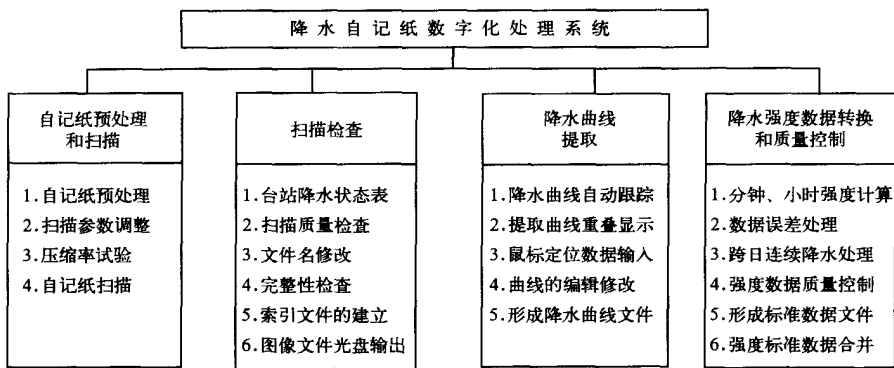


图 1 降水自记纸数字化处理系统结构

② 数据流 降水自记纸经过降水自记纸数字化处理的各阶段,生成各种标准数据文件或中间数据文件,其数据流和数据产品的生成如图 2 所示。

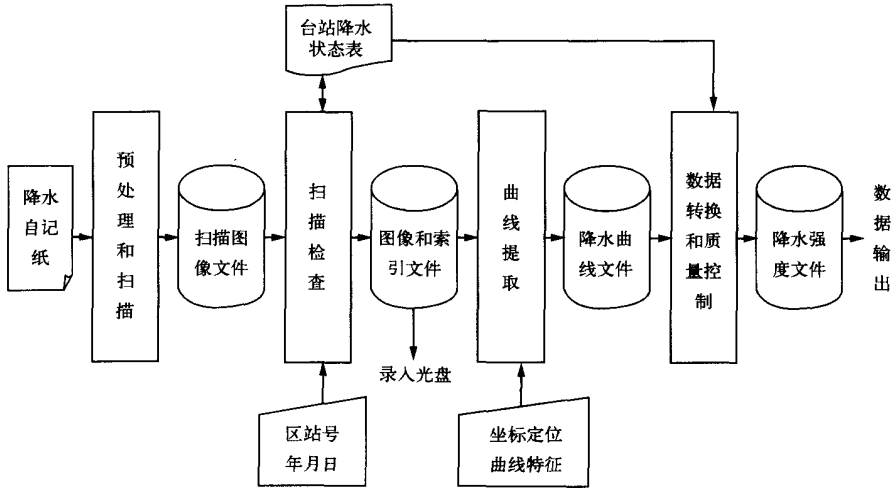


图 2 降水自记纸数字化处理系统数据流和数据产品

2 降水自记纸数字化处理系统功能

2.1 降水自记纸预处理和扫描

对降水自记纸进行扫描输入和数字化处理之前,首先要对降水自记纸进行一系列技术处理,为降水自记纸图像扫描和图形数字化的准确和顺利进行做好准备。这种技术处理,称为降水自记纸预处理。降水自记纸预处理的主要内容有:降水自记纸顺序清理,与降水有关的各种信息检查标注,包括区站号、时间、异常曲线的雨量等。

降水自记纸数字化处理系统的扫描功能是开发利用的 EPSON / GT/10000 + 扫描仪的软件功能。经试验和试用,对于降水自记纸扫描的基本技术参数定为:① 鉴于降水曲线自身精度的限制,扫描分辨率一般取 150 dpi;② 扫描图像采用 JPG 图像压缩文件方式保存。在基本保证 JPG 图像质量的前提下,其压缩率在 95 % 左右。

按规定的操作要求进行扫描,扫描图像清晰,曲线与底色差别明显。

2.2 扫描检查

“扫描检查”要完成对降水自记纸扫描质量、完整性的检查和扫描图像文件的光盘输出。主要内容有:① 建立台站降水状态表,其目的是用于对自记纸降水情况和数字化处理结果的对比,以保证降水自记纸数字化处理的质量,包括区站号、站名、经度、纬度、台站高度和逐日降水量信息;② 扫描图像质量检查,提供图像显示、水平、垂直校准线、多比例尺图像缩放功能;③ 降水自记纸区站号、日期的输入,自动完成图像文件命名;④ 建立无降水和缺测文件(空文件);⑤ 通过文件映像方式进行降水自记纸的完整性(时间连续无缺漏)检查;⑥ 输入文字注释建立备注文件,用于记录降水自记纸的基本情况(如冬季不观测时间)和异常情况;⑦ 建立图像文件的索引文件,将图像文件输出到光盘。

图像文件将上报归档,是系统完成的工作目标之一。

2.3 降水曲线提取

降水曲线提取是指把降水曲线从自记纸彩色扫描图像中分辨出来,并以时间和降水量坐标的方式表征。降水曲线提取采用人机交互方式,自动进行曲线跟踪,完成降水曲线的提取,实现由降水自记纸彩色扫描图像文件到降水曲线数据文件的转换。

2.3.1 降水曲线提取(自动识别和人机交互)功能

降水曲线提取是降水自记纸数字化处理的核心部分,其自动识别和人机交互的能力是决定降水自记纸数字化处理的精度和影响降水自记纸数字化处理工作量的关键因素。本程序设计了较为高效的自动识别技术和较为完备的面对各种复杂情况的人机交互功能,包括:

① 首先设定降水、时间分辨率和跟踪方法(有记忆功能),确定并输入降水曲线的起止时间,然后进行降水曲线的自动跟踪识别,并在降水自记纸彩色扫描图像上重叠显示自动跟踪的曲线。可以适应多种时制(北京时、地方标准时、地方时)、类型(虹吸式、翻斗式)的自记纸,采用多种跟踪方法进行降水曲线的跟踪。

② 为提高曲线自动跟踪效果,设计了如下功能:曲线跟踪色彩阈值的调节,和根据所取阈值生成二值化图像的同步显示功能;消除降水自记纸上全日无降水直线的功能。

③ 对于不能一次性有效完成自动跟踪的情况,提供多种曲线跟踪的交互修改功能:对于降水曲线密集的超强降水的跟踪采用虹吸线端点连接的特殊方式;从某一位置重新开始的补跟踪功能;跟踪曲线的点、线修改功能:单点、多点、跟踪线段的删除,跟踪点的增加、移动;钟筒停走,时间异常处理的功能;曲线上升异常,输入降水量数据或连线功能;翻斗式降水曲线异常跳大格现象的处理。

④ 为保证曲线提取质量,程序提供如下基本功能:当发现图像由于扫描时放置不正出现较明显的歪斜时,进行倾斜旋转校正;虹吸下降线不正(水平歪斜),一般可以认为是钟轴倾斜,进行水平(时间)订正。

⑤ 其他辅助功能:降水自记纸扫描图上时间或降水值的坐标网格显示功能,以方便对“异常”情况的处理和数字化结果的检查或查看有关分辨率的设置是否正确;跟踪数据显示功能;光标精确定位(鼠标微调)功能;跟踪线点的平滑功能,调整曲线跟踪中的局部跳跃波动。

2.3.2 曲线数据存储功能

保存曲线提取后的时间-降水量坐标数据,形成自记纸的降水曲线文件。

2.3.3 检查曲线提取效果的功能

① 显示降水自记纸的曲线提取统计的降水总量和小时降水量,同时显示台站降水状态库的12 h降水量和小时降水量(若有相应的A6 A7文件),提供对照检查。

② 对已经完成的降水曲线文件数据在原图像文件上进行曲线跟踪效果回放,以便检查降水曲线提取的质量。

2.4 降水强度数据转换和质量控制

降水强度数据转换和质量控制是将降水曲线数据转换成降水分钟强度数据,进行数据质量控制,并形成降水强度标准数据文件。

2.4.1 降水分钟强度数据转换和合并

① 降水曲线数据的检查:对降水曲线数据格式的检查;对降水虹吸状态的检查;对相邻日时间是否重叠的检查;数据完整性(日期连续)的检查。

② 降水曲线数据向降水分钟强度数据的转换:把以降水曲线为界限的时间-降水量坐标数据转换为以北京时间 20:00 为日界的每分钟降水量的数据;进行虹吸订正处理、累计进位和数据平滑技术处理。

③ 对降水分钟强度数据的质量检查:进行自记降水分钟强度累计日降水量与雨量筒观测日降水量的比较,有疑问输出相应的信息恭请用户查证核实该数据的正确性;提供疑问数据的文件定位功能。

④ 降水分钟强度数据文件的纠错和合并:为修正降水分钟强度数据文件中的错误数据,提供用新的降水分钟强度数据文件对原降水分钟强度数据文件中错误数据的纠错覆盖功能;具有时间连续检查功能,日期连续的两个降水分钟强度数据文件的合并功能。

⑤ 降水分钟强度数据文件检索:从降水分钟强度数据文件中检索所需的降水分钟强度数据文件。

2.4.2 降水分钟强度标准数据文件的建立

从降水分钟强度数据文件按整年建立降水分钟强度标准数据文件。

2.4.3 降水小时强度的统计和降水小时强度标准数据文件的建立

根据降水分钟强度标准数据文件,按正点统计每小时的降水量,形成降水小时强度标准数据文件。

3 系统设计中的代表性技术

3.1 降水自记纸扫描图形的数字化精度分析和扫描分辨率

扫描图形数据属离散型点阵数据,其精度与分辨率有直接关系。以精度最高的 21 号自记纸为例,降水曲线的宽度在 0.3 mm 以上,相当于降水量 0.03 mm 以上;曲线宽度的变化常常在 0.2 mm 以上,也就是说,曲线自身的精度变化宽度可能造成的曲线跟踪的误差在 0.02 mm 降水量以上;若取扫描分辨率 150 dpi(点/吋)计算,每一像素(点)的间距意味着 0.017 mm 的降水量,因此,扫描分辨率取 150 dpi 已可满足降水自记纸降水曲线的表现精度。

3.2 降水曲线自动跟踪提取

对降水曲线自动跟踪提取,是实现降水曲线数字化的基础和关键,是决定降水曲线图形数字化精度的基础因素。

降水曲线迹线的跟踪提取是降水自记纸图形数字化工作中最重要、最困难、也是工作量最大的工作。降水曲线迹线跟踪提取的自动化能力和操作简易性,是本系统能否广泛投入业务推广应用的 key 部分。

降水自记纸的情况千差万别,表现在:降水曲线迹线粗细不一;降水自记纸底色和网格线颜色深浅不一,降水曲线迹线颜色深浅不一;迹线过密难以分辨;迹线交叉干扰;迹线中断;迹线坐标倾斜;自记纸上污迹干扰;自记纸迹线上升异常;自记纸时间坐标异常等等。

因此降水曲线迹线的跟踪提取必须解决:迹线点特征的有效识别,迹线点中断的有效跟踪,迹线中断后的继续跟踪,密集迹线的有效跟踪,跟踪曲线的局部修正,异常迹线的特殊处理等等。

3.2.1 降水曲线彩色表征区分

降水曲线彩色表征区分技术是曲线自动跟踪提取关键技术之一,是把降水曲线与底图坐标线的彩色数字表征有效区分开来的技术。

利用彩色扫描降水自记纸底色、网格坐标线和降水曲线色彩的明显差异,寻找其色彩数字表征的特征,是明显区分网格坐标线和降水曲线的关键。从亮度、对比度、色彩平衡等几个方面对降水自记纸彩色图形进行的试验表明:消除降水自记纸原图像色彩中绿和蓝色的程度差异,强化图像色彩中红色的差异(只表现为红色和黑色二值),可以有效的去除网格坐标线的痕迹,曲线的色彩呈现为黑色,底色、坐标色则呈现为红色,二者差异显著。也就是说,曲线与底色、坐标色的灰度差异不如单纯红色的差异明显。用红色单色,可以找到明显区分网格坐标线和降水曲线控制的阈值。因此,可以以红色阈值的有效选取作为降水曲线自动跟踪提取基本手段。

3.2.2 降水曲线迹线跟踪提取

降水曲线迹线的跟踪提取是降水自记纸数字化处理系统最关键、最核心的技术。

迹线跟踪一般采用随时间坐标向前搜索具有迹线点特征的点的方法。困难的是,前方具有迹线点特征的点可能有多个,也可能中断。多点的时候需要选择其真正的迹线代表点,中断无点的时候需要寻求新的迹线点。因此,对迹线点的确认必须采用随时间坐标向前延伸搜索的方法。为了达到对迹线点的有效跟踪,进行了多次研究试验,设计了一种“最优法”的跟踪方法,对迹线的跟踪采用四分之一圆多点搜索法,即以当前点为原点,在时间增加和雨量值增加的象限内在5个不同半径的圆周上搜索黑点,每个圆弧上以黑度(是指某个点及其周围8个点色值低于阈值的点的个数,变化范围为0~9)最大且大于3的点作为线迹的候选点并进行合理性检查,去除不合理变化的点。这样一次可以得到1~5个迹线点,它可以跨越10 min或者说0.3 mm的断线继续跟踪。当没有找到候选点时,判断是否为虹吸点(虹吸式自记纸),如果是,则搜索虹吸后的零毫米降水位置再继续搜索,否则停止搜索。

一般情况下,我们将黑度 $> VX$ (通常设 $VX=3$)的点作为迹线的点,但由于阈值的设置有一定的随意性,可能会有一段“迹线”很细(黑白显示可以看到),黑度只能达到3,这样就无法正确识别迹线。这时系统会自动作如下处理:当搜索的黑度值最大的5个点的黑度值都小于4时,临时改变 $VX=2$,即将黑度=3的点作为迹线的点,之后又恢复原设置。

对于强降水(如20 mm/10 min),由于在降水坐标5 mm上下的迹线已经连在一起了,显然无法分得清楚,故采用搜索虹吸点及虹吸后零点的方法得到迹线点。

3.2.3 降水曲线迹线跟踪提取的辅助技术

为提高降水曲线迹线跟踪提取的效果,采用了多种辅助技术,如:

① 图像旋转技术 由于扫描时自记纸放置不正,需要进行图像垂直、水平的校正(任意角度旋转)。为了获取旋转的角度,借鉴了OCR的处理方法,即用鼠标拉一条线,这条线与水平线或垂直线的角度较小的一个就作为图像旋转的角度。

② 图像二值化和滚动条改变阈值、图像同步显示技术 图像二值化的处理是将色值大于阈值的像素点显示为黑点,色值小于阈值的像素点显示为白点。为了方便操作,除了键盘直接修改阈值的功能外,设计了移动滚动条来改变阈值、同步显示相应的二值化图像的功能。并设计了迭加显示和黑白显示两种技术功能,设计了局部二值化和全图二值化两种功能。

③ 鼠标定位、键盘方向键微量移动修正技术。

3.3 图形数字化数据处理技术

提取的降水曲线数据是用纵坐标(降水量)和横坐标(时间)表现的,必须转换成降水分钟强度存储和应用。数据的转换必须遵循降水总量不变的基本原则。针对降水曲线数据的特点,系统采用了如下数据处理技术。

(1) 尾数累计进位及其误差

降水分钟强度数据的计算,是将降水曲线数据两点之间降水量平均分配到相应的分钟时间段上,跨点的分钟时间段降水量,用相应的时间段降水量合计。为了保证降水总量不变,统计值尾数不能采用通常的四舍五入或尾数进、舍的办法取舍,而采用尾数累计进位的办法。统计中发现尾数累计进位有时间延后的现象,是由于两个数除不尽,在运算过程中,可能出现极小的数值误差,影响进位。采用微量增值的办法,解决了计算偏差的问题。

(2) 二项式滑动平均法

降水曲线自动跟踪提取的曲线数据,往往表现出分钟降水强度的振荡式波动变化特征。这种振荡式波动变化,大多不是降水真实情况的反映,而是因为降水曲线自身表征(线条粗细、颜色深浅)的变化和其它痕迹的干扰,造成降水曲线跟踪提取数据的偏差。为了减少这种偏差,按降水时间段对分钟强度数据进行二项式滑动平均的技术处理^[2]。二项式滑动平均技术具有保持降水总量不变、保持降水时间段不变、在降水曲线宽度变化的一定范围内平滑的特点。由于无法明确分辨实况和曲线提取偏差的界限,经二项式滑动平均的降水分钟强度数据,只是在一定程度上减少降水曲线跟踪提取数据的偏差,但无法消除这种偏差。

3.4 同步定位显示技术

系统在提供适应各种情况的较完善的多种功能的同时,为了方便用户使用,采用同步定位显示技术,使用户在面对不同对象时能自动定位同步显示进行作业。例如:①对降水曲线数据检查发现有可疑数据,可以直接打开相应的降水曲线数据文件,并指向可疑数据;②进行降水曲线跟踪,可以同步显示跟踪数据,可以直接删除跟踪数据,同步显示跟踪效果。

4 小 结

在对我国降水、暴雨的研究中,确定时段的降水强度及其分布特征是其最基本的、也是最重要的内容。作为基础资料,过去只有 10 min 最大和 1 h 最大降水强度,无法满足对降水和暴雨深入研究的需要。

降水自记纸彩色扫描数字化处理系统实现了由降水自记纸彩色图像形成每分钟降水量的数字化基础资料,并最终形成长年代每分钟降水量长序列资料的目标。在此基础上,

可以统计产生任意时段的降水强度及其分布特征;可以滑动统计任意时段的最大降水强度。这些资料,对农业、城市、电讯、交通、水利等领域的规划和建设,对防灾减灾策略的制定提供了定量的数字依据,具有重要的意义。

降水自记纸彩色扫描数字化处理系统,于 2003 年开始在全国各省市推广应用。根据应用的情况,我们对系统功能设计又有所改进。目前,我国各省都至少完成了 2 个站的降水自记纸的数字化工作;部分省市已将数字化数据运用到研究和业务领域中去,发挥了效益,例如,重庆市气象局将降水自记纸的数字化数据计算不同时段降水强度用于防治地质灾害的规划设计;广东省气象局已将降水自记纸的数字化数据用于深圳、佛山、清源等城市规划设计。

参 考 文 献

- 1 孙立潭,赵殿军,王铁.气象图形图像处理与应用.北京:国防工业出版社,2002.1~66.
- 2 王伯民.彩色扫描图形数字化处理技术的研究——气象历史档案拯救技术探索之一.应用气象学报,2003,14(6): 763~768.

THE COLOR SCANNING DIGITIZING PROCESSING SYSTEM OF PRECIPITATION AUTOGRAPHIC RECORD PAPER

Wang Bomin¹⁾ Lu Yongping²⁾ Zhang Qiang¹⁾

¹⁾ (National Meteorological Center, Beijing 100081)

²⁾ (Meteorological Bureau of Guangdong, Guangzhou 510080)

Abstract

The purpose of the color scanning digitizing processing system of the precipitation autographic record paper is to realize converting data from color scanning graph to the precipitation intensity per minute. The design thought, framework and functions of the computer processing system are introduced. The master functions of this system include the pretreatment of the precipitation autographic record paper, paper scanning, scan checking-up, precipitation curve collecting and precipitation intensity data conversion. The key technologies in system design causing the reader to all-round know the main content and the characters of this system are also expounded. The representative technologies include the precision analysis and the scan resolution of digitizing the scanning graph, the technique of automatic tracking and collecting the precipitation curves, the technique of data processing of digitizing graph, the synchronous orientation display technique etc. The color scanning digitizing processing system of precipitation autographic record paper has been applied to operations in provinces.

Key words: Color scanning Digitizing graph system Precipitation recording graph