

# 温度自记纸迹线数字化提取软件设计与实现

张茜茹<sup>1,2</sup>, 陈益玲<sup>1,2\*</sup>, 李长军<sup>1,2</sup>, 张军<sup>3</sup>

<sup>1</sup>山东省气象防灾减灾重点实验室, 山东 济南

<sup>2</sup>山东省气象信息中心, 山东 济南

<sup>3</sup>重庆众仁科技有限公司, 重庆

收稿日期: 2022年8月20日; 录用日期: 2022年9月19日; 发布日期: 2022年9月28日

## 摘要

温度自记纸是客观记录温度的十分宝贵的历史气象资料, 记载了分钟级的温度变化特征。为了满足气象防灾减灾以及气象信息化业务的需要, 基于图像处理技术, 采用图像矫正、器差订正多种检查矫正方法, 研发了温度自记纸迹线数字化提取软件, 实现了对温度自记纸图像的认识提取和质量控制, 为最终形成长时间序列、高分辨率的温度自记纸数据提供了技术支持。

## 关键词

温度自记纸, 迹线识别, 数据提取, 检查矫正

# Design and Implementation of Digital Extraction Software for Temperature Self-Recording Paper Trace

Qianru Zhang<sup>1,2</sup>, Yiling Chen<sup>1,2\*</sup>, Changjun Li<sup>1,2</sup>, Jun Zhang<sup>3</sup>

<sup>1</sup>Key Laboratory for Meteorological Disaster Prevention and Mitigation of Shandong, Jinan Shandong

<sup>2</sup>Shandong Meteorological Information Center, Jinan Shandong

<sup>3</sup>Chongqing Zhongren Technology Co., Ltd., Chongqing

Received: Aug. 20<sup>th</sup>, 2022; accepted: Sep. 19<sup>th</sup>, 2022; published: Sep. 28<sup>th</sup>, 2022

## Abstract

Temperature self-recording paper is a very valuable historical meteorological data that objective-

\*通讯作者。

文章引用: 张茜茹, 陈益玲, 李长军, 张军. 温度自记纸迹线数字化提取软件设计与实现[J]. 计算机科学与应用, 2022, 12(9): 2195-2202. DOI: 10.12677/csa.2022.129223

ly records temperature, and records the minute-level temperature change characteristics. In order to meet the needs of meteorological disaster prevention and mitigation and the meteorological information business, based on image processing technology, multiple inspection and correction methods such as image correction and device error correction are used, and a digital extraction software for temperature self-recording paper trace is developed. This software realizes the identification, extraction and quality control of temperature self-recording paper images and provides technical support for the final formation of long-time series, high-resolution temperature self-recording paper data.

## Keywords

Temperature Self-Recording Paper, Trace Identification, Data Extraction, Inspection and Correction

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

自记纸是客观记录天气现象的十分宝贵的气象资料,我国在 20 世纪 50 年代就开始了自记观测,各省积累了大量的历史风、温、湿、压和降水自记纸,数据资源非常丰富。但由于早期受档案载体及信息技术限制,没有对自记纸进行数字化处理,制约了自记纸数据的利用。随着信息技术的发展,各省份开始对大量的自记纸历史资料进行数字化处理工作[1][2],自记纸数据的存储和利用已成为现实。

随着计算机图像识别技术的发展,各省份先后完成了降水自记纸、EL 型电接风自记纸迹线数字化提取[3],但是温度自记纸提取工作尚未开展。目前,降水自记纸、EL 型电接风自记纸迹线提取软件对轻微扭曲、歪斜的图像仍采用人工重扫的方法,尚未实现对图像的自动订正,同时已有软件未对自记仪器本身的误差进行订正,不能满足温度自记纸的智能化、高质量的提取要求,因此研发温度自记纸迹线数字化提取软件并优化检验矫正技术十分必要。按照中国气象局下达的温度自记纸数字化有关规定,山东省气象局承担了温度自记纸迹线数字化提取与质量控制技术研发任务。基于已有的自记纸提取技术基础,山东省研发了图像矫正、器差订正等检验矫正功能,实现了温度自记纸数字化工作半自动化、智能化,极大地提高数字化工作的生产效率,节约人力成本、提高生产产值与质量。本文从温度自记纸迹线数字化提取软件的功能设计、关键技术角度出发,介绍了分钟级、小时级温度自记纸数据的提取流程。

## 2. 软件设计

### 2.1. 软件结构设计

温度自记纸迹线提取软件基于 C/S 结构,采用文本型数据库 SQLite,兼容 Windows XP、Windows 7 和 Windows10 操作系统,实现对温度自记纸历史档案迹线的数字化跟踪提取。本软件主要分为数据批量提取、数据矫正修改、数据质量检查与数据集生成三大模块,软件结构如图 1 所示:

### 2.2. 软件流程设计

温度自记纸迹线数字化提取流程主要分为以下 4 个步骤:

- 1) 温度自记纸图像文件整理。软件自动对扫描图像进行可读性和连续性检查,对缺漏、不合格图像

文件进行补扫。

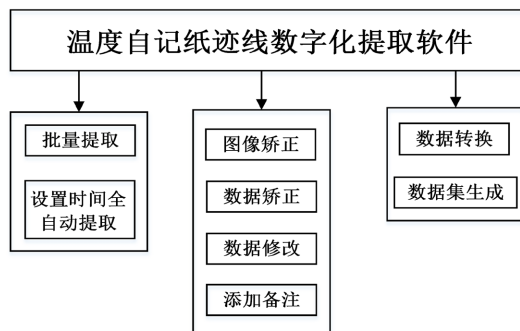


Figure 1. Software structure design diagram

图 1. 软件结构设计图

2) 温度迹线提取。首先需要对导入的图像文件的基础配置信息进行设置, 包括设置迹线和中轴线开始结束时间、业务值区间等, 然后对迹线进行自动识别提取。

3) 温度迹线检查修正。对批量提取的迹线数据进行回放检查对比, 未通过检查的通过图像矫正、手动修改、器差订正等方式修改提取有误的迹线数据, 并针对不同情况添加备注。

4) 数据转换与产品生成。通过数据质量控制的迹线坐标文件转换为分钟数据(T1)文件、标准数据文件(T01)以及标准小时数据文件(T60)。软件总体流程如图 2 所示:

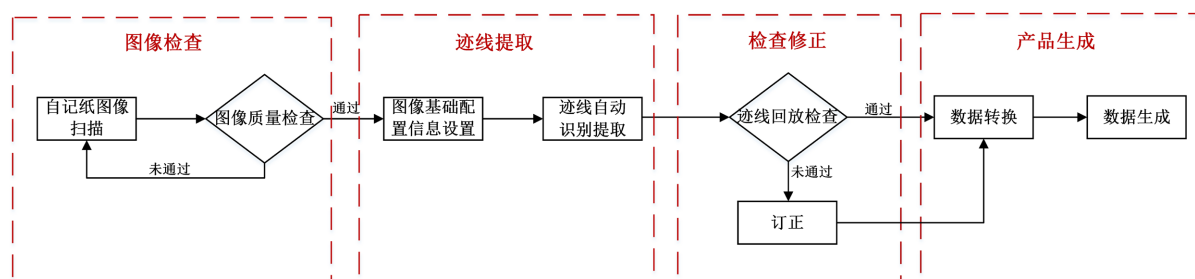


Figure 2. Software process design diagram

图 2. 软件流程设计图

### 3. 软件功能

#### 3.1. A 文件导入

A 文件导入模块是将人工观测的定时观测结果导入本地文本数据库, 用于迹线识别数据的质量控制。通过导入 A 文件, 可以获取文件日期、数值、观测时次、数据类型和文件类型等信息(图 3)。

#### 3.2. 数据批量处理

数据批量处理模块实现对每张温度自记纸迹线识别与数据提取, 运行界面如图 4 所示。具体功能如下:

1) 图像导入。导入扫描生成的自记纸图像文件, 软件将自动进行图像文件名、日期连续性检查, 并对图像文件分辨率、倾斜度进行检查, 要求图片分辨率  $\geq 300$  dpi, 图像倾斜度  $\leq 0.5^\circ$  [4]。如果不能满足自记纸图像文件相关要求, 则对该文件进行重扫。

项目	A文件导入	批量处理	检查修正	成品数据	图像矫正	设置	帮助
选择文件	开始	日期	值	类型	方式位	文件类型	几次定值
1	1973-12-01 02:00			TIMING	T0	A0	0200,0800,1400,2000
2	1973-12-01 08:00			TIMING	T0	A0	0200,0800,1400,2000
3	1973-12-01 14:00			TIMING	T0	A0	0200,0800,1400,2000
4	1973-12-01 20:00			TIMING	T0	A0	0200,0800,1400,2000
5	1973-12-01 00:00			MAX	T0	A0	0200,0800,1400,2000
6	1973-12-01 24:00			MIN	T0	A0	0200,0800,1400,2000
7	1973-12-02 02:00			TIMING	T0	A0	0200,0800,1400,2000
8	1973-12-02 08:00			TIMING	T0	A0	0200,0800,1400,2000
9	1973-12-02 14:00			TIMING	T0	A0	0200,0800,1400,2000
10	1973-12-02 20:00			TIMING	T0	A0	0200,0800,1400,2000
11	1973-12-02 00:00			MAX	T0	A0	0200,0800,1400,2000
12	1973-12-02 24:00			MIN	T0	A0	0200,0800,1400,2000
13	1973-12-03 02:00			TIMING	T0	A0	0200,0800,1400,2000
14	1973-12-03 08:00			TIMING	T0	A0	0200,0800,1400,2000
15	1973-12-03 14:00			TIMING	T0	A0	0200,0800,1400,2000
16	1973-12-03 20:00			TIMING	T0	A0	0200,0800,1400,2000
17	1973-12-03 00:00			MAX	T0	A0	0200,0800,1400,2000
18	1973-12-03 24:00			MIN	T0	A0	0200,0800,1400,2000
19	1973-12-04 02:00			TIMING	T0	A0	0200,0800,1400,2000

Figure 3. A-File import module operation renderings  
图 3. A 文件导入模块运行效果图

- 2) 参数设置。通过对基础信息和扩展信息(迹线开始结束时间、中轴线开始结束时间和业务值区间)进行配置,为后续迹线识别与数据提取子系统提供每张温度自记纸的初始输入参数。
- 3) 图像提取。基于获取潜在迹线点和迹线形态特征去噪等图像处理技术[5] [6],实现温度自记纸批量自动识别和数据提取,提取后形成迹线坐标 TXT 文件。软件批量提取完成后,页面显示提取正常、失败、丢失、重复、缺测和备注文件数量,通过“异常信息导出”功能导出数据中存在的异常日期数据。

Figure 4. Data batch processing module operation renderings  
图 4. 数据批量处理模块运行效果图

### 3.3. 检查修正

采用温度迹线回放方式，对批量提取的迹线数据进行检查，运行界面如图 5 所示。主要检查修正内容如下：

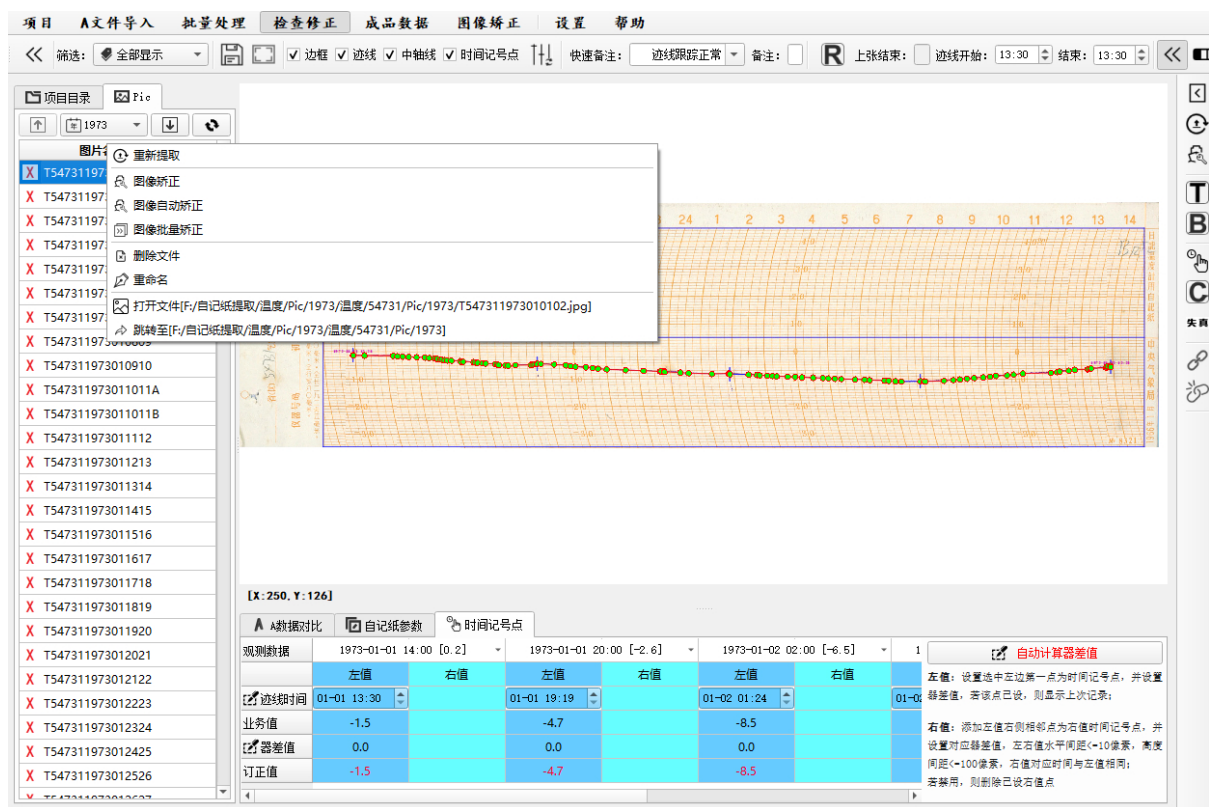


Figure 5. Check and correction module operation renderings

图 5. 检查修正模块运行效果图

1) A 数据对比。即将导入的 A 整点数据、极值数据与迹线提取值进行对比，并根据气象规范设定误差等级，用于后续迹线数据的检查修正。

2) 自记纸参数设置。输入图像文件的基础配置信息，主要包括中轴线开始、结束时间，时间分辨率和曲率半径等基础信息设置。

3) 图像矫正。基于降水、风自记纸的扫描经验，温度自记纸迹线提取软件研发了图像矫正功能。通过图像矫正功能可以对所有连续、渐近的轻微形变进行处理，主要处理图像倾斜、扭曲变形、挤压变形等异常图像，符合自记纸由于长时间累积形变的应用场景。

4) 迹线矫正。对自动提取过程中温度迹线跟踪效果不好或需要进行数据订正、备注录入等情况进行修改，常规操作主要包括迹线点增、删、改、时间记号点以及仪器异常等备注信息录入。

5) 器差订正。在温度自记纸迹线提取软件中新增了器差订正功能，对迹线各时间点读数进行器差订正，消除自记仪器本身的误差。器差值是指定时观测与迹线值之间的差值，在本软件中可以手动设置器差值，也可以点击“自动计算机器差值”选项，软件自动计算。设置完器差值后，软件进行自动订正，订正后“时间记号点”处的迹线值应与定时观测值保持一致；此外，求出两定时观测器差值的差值，需要将其线性分配到两定时观测时间记号之间的各时间点，实现对迹线各时间点读数的订正。但需要注意的



是, 若在时间记号点位置出现跳跃迹线, 则需要通过设置时间记号点左值器差和右值器差对业务数据分段订正, 使时间记号点左右的业务值均等于对应整点的定时观测业务值, 保证转化迹线数据的准确性和连续平滑性。

### 3.4. 成品数据

该模块包含两个子模块: 一是数据生成模块, 二是数据合并模块, 功能界面如图 6 所示。数据生成指将数据进行质量检查, 如果没有明显矛盾数据, 则将所提取的迹线数据批量转化为以月为单位的标准数据, 包括小时数据和分钟数据; 数据合并指如果存在多人加工一个站的情况, 需要将多个加工数据进行合并操作。数据文件命名格式如下:

自记纸小时数据记录格式: ThIIiii-yyyymm.txt; 自记纸分钟数据记录格式: TmIIiii-yyyymm.txt。其中, T 代表温度, h 代表小时数据, m 代表分钟数据, IIiii 表示区站号, yyyy 表示年份, mm 为月份(位数不足, 高位补“0”)。

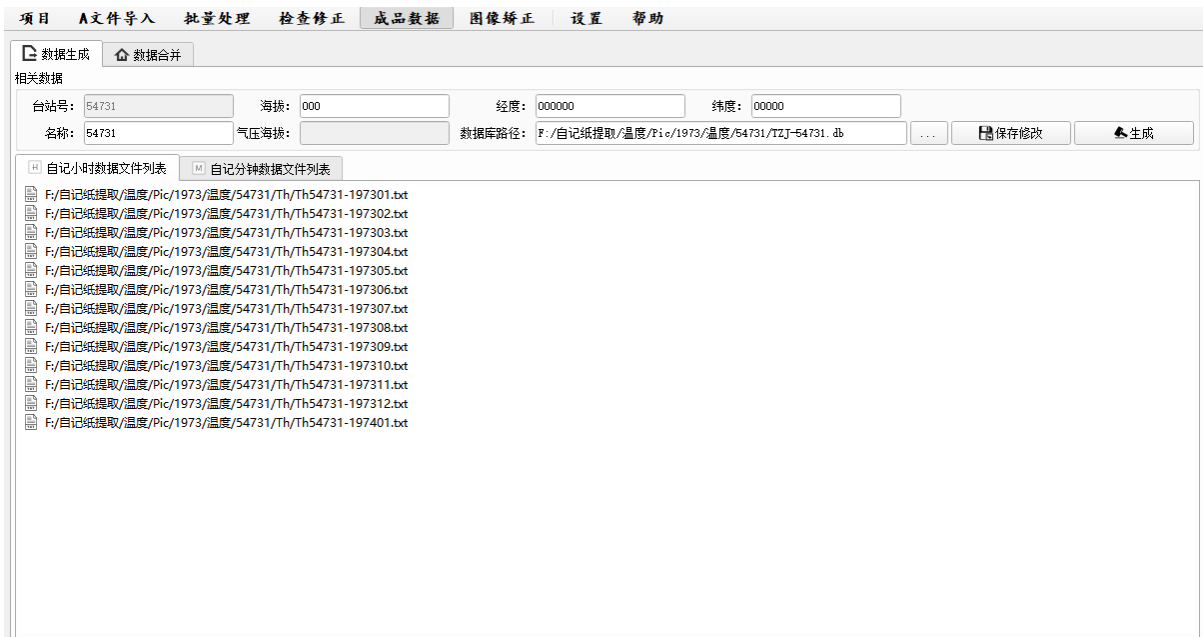


Figure 6. Finished product data module operation renderings

图 6. 成品数据模块运行效果图

## 4. 关键技术

### 4.1. 图像提取和迹线识别

温度自记纸图像提取主要包括边框提取和迹线提取两部分, 处理过程与关键技术如下:

1) 边框数据与迹线数据的分离。计算机图像处理中, 图像的色彩是通过红(R)、绿(G)、蓝(B) 3 个颜色通道的变化以及相互叠加来得到各种颜色[7]。根据统计(统计前首先去除图中绝大多数白色背景点), 在温度自记纸中, R 值较大的像素值和 B 值较小的像素值大致对应了表格线的像素点, 反之则对应迹线类点, 包括少量噪声点。因此, 采用 R、B 通道的差值  $RB (RB = R - B)$  作为区分边框数据和迹线数据的阈值条件。

2) 边框数据提取。根据表格线呈水平或者竖直分布特征, 采用水平和竖直投影方式确定上下、左右

边框的位置。

3) 迹线提取。对 RB 值进行升序排序, 取前 5% 的像素点作为潜在迹线点进行提取, 提取结果如图 7 所示。提取后, 检测二值图像上所有连通域的形态特征, 包括周长面积比、长宽比、空心率等, 判断连通域是否为字体、噪声等, 然后消除干扰连通体, 将剩下连通域拟合为平滑曲线, 即为温度迹线。

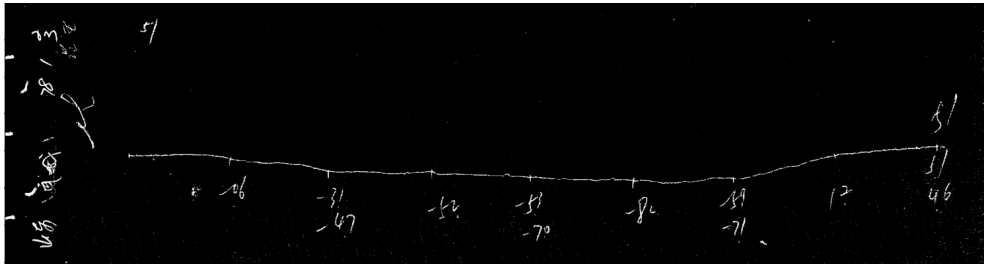


Figure 7. Extraction result of potential trace points  
图 7. 潜在迹线点提取结果

#### 4.2. 图像矫正

图像矫正是根据轻微变形图像的渐近性、连续性特征, 将倾斜、扭曲的异常图像矫正为正常图像的过程。由于温度自记纸档案保存时间久远, 部分档案有不同程度的挤压和扭曲变形, 使得扫描图像存在轻微形变, 从而使边框坐标系获取数据存在误差或者迹线数据提取失败。为了使温度迹线数字化工作高效有序的开展, 软件提供了对异常图像的矫正功能。对于整个图像而言, 图像边框的倾斜程度决定了边框坐标系的精确程度, 而边框坐标系决定了时间记号点对应的的时间, 进而决定了迹线数据转化的精度, 所以对倾斜、变形的边框线进行矫正, 有利于提高迹线数据的提取质量。

本文以水平矫正为例, 介绍了图像矫正的原理。如图 8(a)所示, 自记纸中段与蓝色水平线之间的距离较小, 而两端较大, 扫描图像中段存在下垂变形。通过软件手动跟踪变形自记纸上下边框, 根据轻微变形的连续性和渐近性, 对上下边界点中段进行线性插值, 确定每列需要插值和平移的距离, 矫正结果如图 8(b)所示。

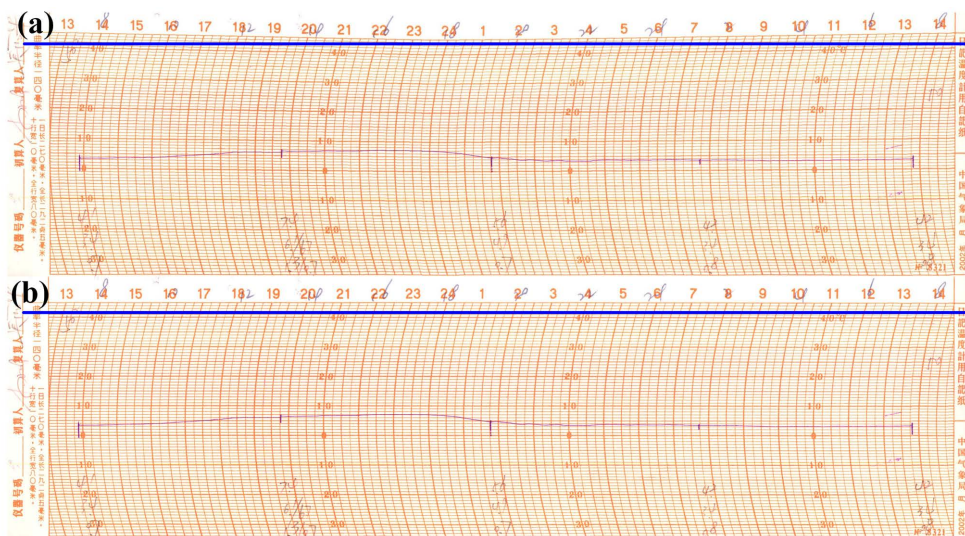


Figure 8. Slightly deformed image example (a) and image corrected effect (b)  
图 8. 轻微变形图像示例(a)与图像订正后的效果(b)

### 4.3. 器差订正

由于自记仪器本身存在一定误差,所以对迹线数据进行器差订正十分必要。器差订正依赖时间记号点,每个时间记号点默认器差为 0,可以手动设置,也可通过软件自动计算,迹线对应所有分钟数据由其相邻的左右时间记号点器差线性决定,即先求出两定时观测器差值的差值,然后将其线性分配到两定时观测时间记号之间的各时间点。器差订正按公式(1)进行:

$$P_i = U_i + D_0 + \frac{T_i - T_0}{T_m - T_0} \times (D_m - D_0) \quad (1)$$

式中  $P_i$  为待求时间点订正要素值;  $U_i$  为  $i$  时间温度数值;  $D_0$  为前一定时观测做记号右侧器差值;  $T_i$  为带求时间;  $T_0$  为前一定时观测做记号时间;  $T_m$  为后一定时观测做记号时间;  $D_m$  为后一定时观测做记号左侧器差值。

## 5. 总结

本文针对温度自记纸的数字化问题,基于图像识别技术,采用图像矫正、器差订正多种检查矫正方法,研发了温度自记纸迹线数字化提取软件。通过对轻微扭曲、歪斜图像的自动订正,以及对自记仪器本身误差的订正,极大地提高了数字化工作的生产效率和迹线数据的准确率。该软件的研发为全国长序列温度数据集的建立提供了技术支持,加强了对气象数字资源管理,持续丰富了气象档案资源,全面提高了气象业务信息化水平。

## 基金项目

山洪地质灾害防治气象保障工程 2022 年建设项目-1 标段:档案数据资源提取和利用——达因风自记纸、温度自记纸数据提取建设(HSZB-2022-GK024)。

## 参考文献

- [1] 李长军,黄少平,陈益玲.降水自记纸数字化成果质量控制与评估技术[M].北京:气象出版社,2019:1-2.
- [2] 李亚丽,黄少平,鞠晓慧.基于边缘检测法的风自记纸图像数字化[J].计算机技术与发展,2022,32(1):160-164+169.
- [3] 王钰,黄少平.EL型电接风自记纸数字化成果质量控制软件应用评估[J].江西科学,2022,40(4):772-775.
- [4] 黄少平,鞠晓慧,李长军,等.QX/T626-202.气压、气温、相对湿度自记纸记录数字化[S].北京:气象出版社,2021.
- [5] 崔巍.气象观测自记纸图形数字化处理技术的应用[J].通讯世界,2016(18):275-276.
- [6] 岑瑶.图像处理技术在实现气压自记纸数字化方面的应用[J].电子技术与软件工程,2020(2):133-134.
- [7] 赵晓莉,苑跃,黄晓龙,等.EL型电接风向风速自记纸数字化风向识别方法研究[J].中低纬山地气象,2019,43(4):83-86.