

DOI: 10.13476/j.cnki.nsbdtqk.2020.0100

郑森,顾冲时,邵晨飞. 基于图像处理技术的大坝监测数据粗差识别[J]. 南水北调与水利科技(中英文), 2020, 18(5): 123-129.
ZHENG S, GU C S, SHAO C F. Recognition of gross error of dam monitoring data based on image processing technology[J].
South-to-North Water Transfers and Water Science & Technology, 2020, 18(5): 123-129. (in Chinese)

基于图像处理技术的大坝监测数据粗差识别

郑森, 顾冲时, 邵晨飞

(河海大学 水利水电学院, 南京 210098)

摘要:为实现大坝长久稳定的安全服役,避免监测资料中的粗差对大坝安全监测结果产生影响,需要对监测数据中的粗差进行剔除。由于目前的粗差识别方法依旧会造成粗差漏判、误判情况的发生,通过模仿人工识别数据粗差的过程,运用程序设计语言,提出一种基于图像处理技术的自动化粗差识别方法。首先对依据监测数据绘制出的散点图进行高斯模糊和二值化处理,再提取主要趋势线,最后识别出监测数据中的粗差点并进行剔除。选取某实际工程大坝监测资料,运用该方法对其进行粗差识别,并与传统 3σ 识别准则的粗差识别效果进行对比。算例结果表明:该方法对数据粗差的识别效果更加显著,避免了粗差漏判情况的发生,对粗差的剔除更彻底;利用该方法识别后得到的统计模型复相关系数为0.999,标准差为0.192,模型精度更高,也更符合工程实际情况。因此,该方法具有一定的工程应用前景和实用价值。

关键词:监测数据;高斯模糊;图像二值化;粗差识别;统计模型

中图分类号:TV698 文献标志码:A 开放科学(资源服务)标志码(OSID):



Recognition of gross error of dam monitoring data based on image processing technology

ZHENG Sen, GU Chongshi, SHAO Chenfei

(Water Conservancy and Hydropower Engineering, Hohai University, Nanjing 210098, China)

Abstract: In order to realize the long-term and stable service of the dam and avoid the influence of gross errors in the monitoring data on the dam safety monitoring results, it is necessary to eliminate the gross errors in the monitoring data. Because the current gross error recognition method can still cause the gross error to be missed or misjudged, an automatic gross error recognition method based on image processing technology by imitating the process of manual data gross error recognition and using a programming language. The scatter map drawn according to the monitoring data is processed by Gaussian blur and binarization, the main trend line is extracted, and the gross error in the monitoring data is identified and eliminated. The monitoring data of a real dam are selected, the gross error is identified, and the results are compared with those of the traditional 3 identification criteria. The results show that: the recognition effect of the method is more significant, the applied method avoids the false negatives of gross errors and eliminates gross errors more thoroughly, in addition, the complex correlation coefficient of the statistical model obtained by the method is 0.999, while the standard deviation is 0.192, which shows that the accuracy of the model is higher and the model is more in line with the actual situation of the project. Therefore, the method has a certain engineering application prospect and practical value.

Key words: monitoring data; Gaussian blur; image binarization; gross error recognition; statistical model

收稿日期:2019-11-20 修回日期:2020-02-04 网络出版时间:2020-03-16

网络出版地址:<http://kns.cnki.net/kcms/detail/13.1334.tv.20200313.1504.008.html>

基金项目:国家自然科学基金(51739003)

作者简介:郑森(1997—),女,山西洪洞人,主要从事水工结构安全监控。E-mail:22921092@qq.com

通信作者:顾冲时(1962—),男,江苏南通人,教授,博士,主要从事水工结构安全监控。E-mail:csu@hhu.edu.cn

随着监测技术、通信技术和电子产业的迅速发展,中国大坝安全监测自动化得到长足的发展,大坝安全监测资料的数据量也越来越庞大^[1-3]。但在大坝变形监控数据中也存在各种异常数据,这些数据在处理时被认定为粗差点,即监测数据中误差绝对值大于 3 倍限定误差值的数值点。因此,需要对监测数据中的粗差点进行准确识别才能正确评价大坝的工作性态^[4-6],以往的粗差识别方法有过程线法、统计检验法和函数模型法^[7]等。程浩然等^[8]提出了一种基于统计学方法处理大坝安全监测中粗差的方法,虽然减少了计算量,但由于引入了函数模型方法,无法应用于变化规律不确定的数据序列;陈向阳等^[9]将小波分析应用于 GPS 动态监测数据粗差探测中,但该方法中的小波基函数的支撑区间有限;邓波等^[10]基于四分点法对大坝变形监控数据进行粗差识别,但该方法只能识别在一定误差范围内的粗差点;景继等^[11]基于抗差最小二乘法提出了水工监测数据粗差探测方法,但该方法也无法应用于变化规律不确定的数据序列;李兴等^[12]提出了基于 M 估计的改进 Pauta 准则的粗差识别方法,该方法面对波动较大的数据序列时粗差识别效果一般。这些方法虽然粗差识别效果较好,但只能针对某一类变化规律的数据序列,当监测数据含有较多离群数据点或数据变化情况复杂时,极易造成粗差漏判情况的发生。

为寻找一种能够应用于变化规律不确定的复杂数据序列的粗差处理方法,避免产生异常值漏判问题,本文模拟人工识别粗差的过程,创造性地提出一种基于高斯模糊和图像二值化处理技术的粗差识别方法,无需建立函数模型,即可应用于变化趋势不确定的数据粗差处理。该方法依靠 Python 程序设计语言,发挥 Python 语言简洁清晰、兼容性高的优点,实现对数据粗差的自动化智能识别,通过该方法与传统 3σ 准则的粗差识别效果进行对比,探讨该方法的可行性、简便性、合理性和适用性。

1 图像高斯模糊与二值化原理

1.1 图像高斯模糊原理

高斯模糊也称高斯平滑,是一种图像模糊滤波器,利用高斯模糊对散点图进行处理,可以凸显连续点,提取主趋势线,减淡粗差点。该图像处理技术已在 Adobe Photoshop、GIMP 以及 Paint.NET 等图像处理软件中得到广泛应用,一般被用来减少图像的噪声和细节。其对图像的处理过程可以理解为高斯分布的卷积核与原始清晰图像进行卷积运算,通

过对图像中某点周围的像素值进行正态分布曲线的记录,最后通过数学加权平均产生成果图像,原始清晰图像像素的值有最大的高斯分布值,故有最大的权重,相邻像素随着距离原始像素越来越远,其权重也越来越小。高斯分布也被称为正态分布,其密度函数为

$$G(x) = \frac{1}{\delta \sqrt{2\pi}} e^{-\frac{1}{2}(x-u)^2/\delta^2} \quad (1)$$

式中: x 为自变量; u 为平均值; δ 为标准差, δ 越大,高斯函数的宽度越宽。一般平均值取零,对式进行二维平面的推广^[20-21]

$$G(x, y) = \frac{1}{2\pi\delta^2} e^{-\frac{1}{2}(x^2+y^2)/\delta^2} \quad (2)$$

式中: x, y 均为自变量,其余参数含义与式(1)一致。依据式(2),可以对空间内任意点进行高斯模糊,权重矩阵表示见表 1。

表 1 高斯模糊权重矩阵

Tab. 1 Gaussian fuzzy weight matrix

(-1,1)	(0,1)	(1,1)
(-1,0)	(0,0)	(1,0)
(-1,-1)	(0,-1)	(1,-1)

1.2 图像二值化原理

图像的二值化处理是指将图像上点的灰度值设为 0 或 255,使整个图像呈现出明显的黑白极化效果。首先将图像灰度进行二值化设置,使得此时图像的集合性质只与像素值为 0 或 255 的点的位置有关,不再涉及像素的多级值,以此简化处理的复杂度,压缩图像的数据量。由于图像灰度一般包括 256 个亮度等级,通过设定合适的阈值,所有灰度大于或等于阈值的像素灰度值设置为 255,否则这些像素点被排除在物体区域外,并将灰度值设置为 0,表示这些像素点不再被考虑在物体之内^[22]。

在散点图经过高斯模糊之后,散点外侧的像素被减淡,连续点由于聚集在一起使整体像素被增强。因此,连续点整体灰度值大于设定阈值,便被全部设定为 255。粗差点由于孤立在外,以单点的形式存在,经高斯模糊后整体被减淡,灰度值小于设定阈值,便被全部设定为 0。经过二值化处理,散点图可以呈现十分明显的分割效果,由连续点形成的主趋势线便被突显出来,粗差点被有效识别与消除。

1.3 散点绘制形式与大小

高斯模糊与二值化图像处理技术都需要综合考虑连续点覆盖情况和粗差点孤立情况,而连续点的覆盖情况与粗差点的孤立情况受绘制点两个主要因

素的影响:一是散点的绘制形式 χ ;二是散点的绘制大小 s 。为得到较好的图像处理效果,本文对散点的绘制形式 χ 和绘制大小 s 进行了选取。

1.3.1 散点绘制形式 χ 的选取

关于数据点绘制形式 χ 的选取,本文选取了3种具有代表性的散点绘制形状:“×”形状、“·”形状和“+”形状。图1为3种绘制形状的4个包含9个像素的连续点覆盖效果,其中红点代表4个连续点

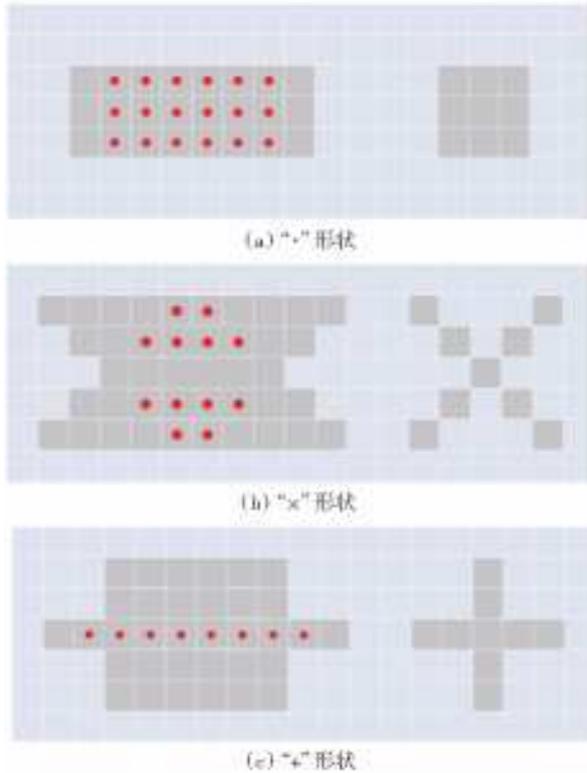


图1 不同形状绘制点堆叠后的覆盖面积
Fig.1 Coverage area of different shape drawing points after stacking

1.3.2 散点大小 s 的选取

关于散点大小 s 的选取,本文选取了5、9、13和17个像素大小的散点进行对比,见图3。“×”形状散点太小便会接近于“·”形状的散点的绘制效果,太大则难以将粗差点孤立出来。

图3为5、9、13和17个像素大小的“·”形状的散点绘制成的散点图经过高斯模糊和二值化后的效果。由图3可知:9个像素大小的散点经高斯模糊和二值化后比其余大小的散点更能突显连续点,减淡粗差点,故选择大小为9个像素的“×”形状的散点绘制散点图。

2 粗差识别程序步骤及设计

2.1 粗差识别程序步骤

在本方法中高斯模糊强度 α 设为2,二值化阈值 β 为150。该粗差识别方法的步骤。

重叠的像素点。由图1可以看出:“×”形状的绘制点和“+”形状的绘制点比实心的“·”形状的连续散点具有更小的重叠度和更大的覆盖面积。图2左侧展示了不同形状4个连续点经过高斯模糊的效果,右侧展示了粗差点经过高斯模糊后的效果,经过对比可以明显看出:“×”形状的散点比其他2种散点更能区分出连续点和粗差点,因此初步选择“×”形状的绘制点绘制数据散点图。

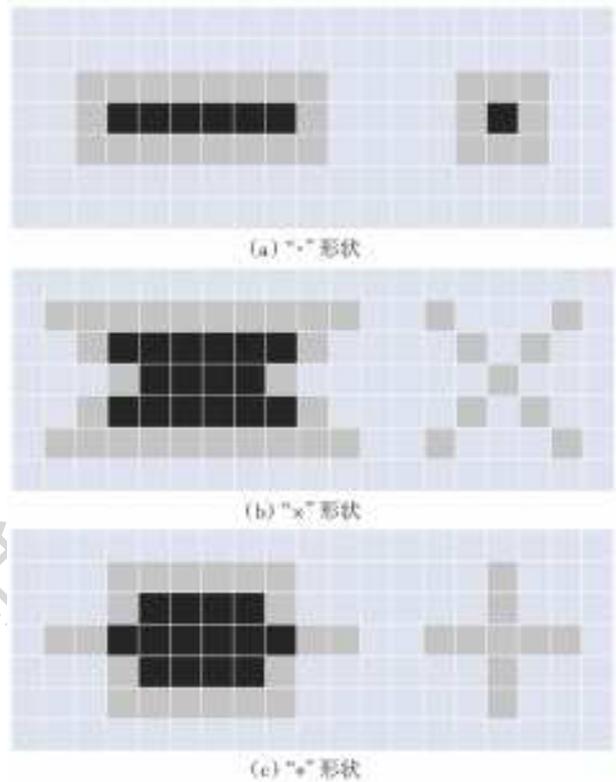


图2 不同形状连续点和粗差点经过高斯模糊后的效果
Fig.2 The effect of continuous points and gross error points with different shapes after Gaussian blur

(1) 依据监测数据绘制散点图,提取监测数据的最大、最小值,根据最值大小设置数据散点图的纵坐标范围。

(2) 对绘制的散点图进行高斯模糊和图像二值化处理,得到主要趋势线。

(3) 认定趋势线上的数据点为连续点并保留其数据值。

(4) 认定处理后被消除的数据点为粗差点并进行剔除。

2.2 基于Python语言的程序设计

Python是一种不受限制,可以跨平台、面向对象的开源编程语言,几乎在任意的操作系统上都能运行。用户能够利用Python脚本语言在ArcGIS软件环境中,因为可以满足数据快速处理的需要,可以被用于具有庞大数据量的大坝监测资料处理

中^[13-16]。NumPy 系统是 Python 的一种开源的数值计算扩展,因为比 Python 自身的嵌套列表结构

在存储和处理大型矩阵方面具有更大的优势,通常被用于图片处理之中^[17-19]。

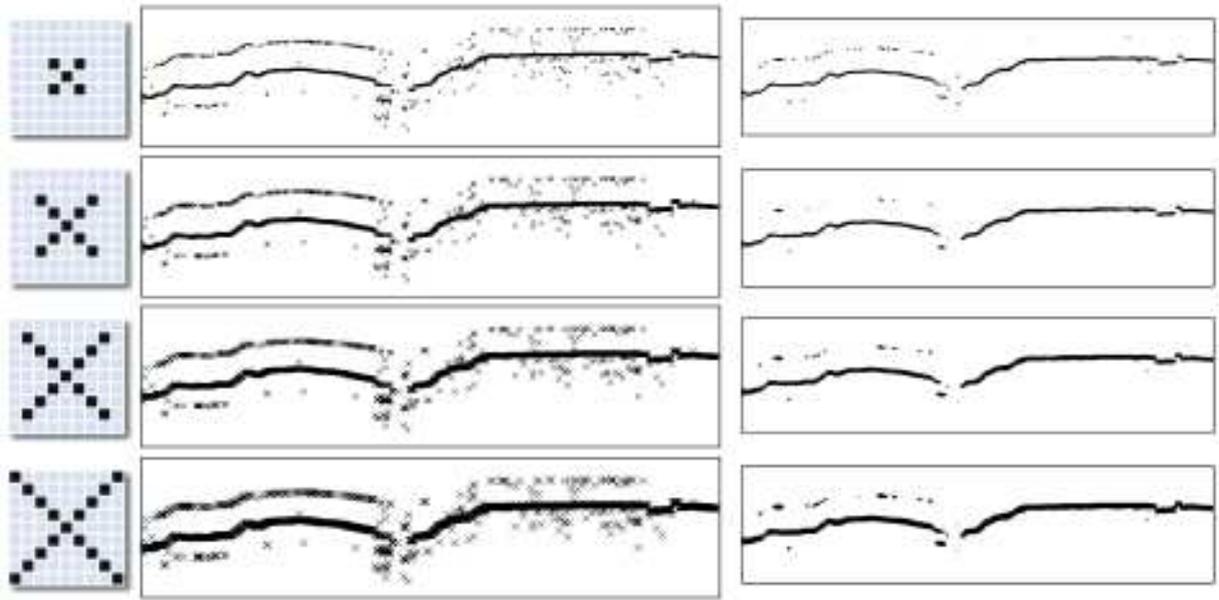


图 3 不同大小绘制点经高斯模糊和二值化后的效果

Fig. 3 The effect of drawing points with different size after Gauss blur and binarization

以下是本文粗差识别程序的实现步骤与关键设计内容。

(1)将目录下 Excel 表中的数据输入,在 Python 软件中导入以下工具包。程序关键设计内容为

```
import sys
sys.path.append('D \位移.xlsx')
import numpy as np
import matplotlib.pyplot as plt
```

(2)进入主程序模块,设置最小数据个数 sjzx、最小覆盖阈值 fgyz、高斯模糊参数 gsmh 和图像二值化阈值 ejyz。程序关键设计内容为

```
sjzx=20
fgyz=80
gsmh=2
ejyz=150
```

(3)创建一张空白图表,绘制出每个点 9 个像素且为“×”形状的散点图,之后直接对散点图进行高斯模糊与二值化处理,得到一条主趋势线。程序关键设计内容为

```
graph = np.zeros((int(len(data)/2)+5,
height+10), 'uint8')+255
if maxx==minn:
ys=height
else:
ys=height/(maxx-minn)
for i in range(len(data)):
```

```
x=int(i/2)+2
y=5+int((float(data[i,1])-minn)*ys)
if y>=5 and y<=height+5:
graph[x,y]=0
graph[x+1,y+1]=0;graph[x+2,y+2]=0
graph[x-1,y+1]=0;graph[x-2,y+2]=0
graph[x+1,y-1]=0;graph[x+2,y-2]=0
graph[x-1,y-1]=0;graph[x-2,y-2]=0
im=filters.gaussian_filter(graph,gsmh);
im=1*(im<=ejyz);
```

(4)当数据点像素过小时,识别为粗差点并剔除,程序关键设计内容为

```
for each in regionprops(label(im)):
if each.area<10:
temp=each.coords
for j in range(len(temp)):
im[temp[j,0],temp[j,1]]=0
lm=label(im)
props=regionprops(lm);n1=len(props)
if n1==0:
break
```

(5)保留有效数据值并输出,粗差识别任务完成,程序关键设计内容为

```
for i in range(len(data)):
if data[i,2]!=1:
d[int(data[i,0])]=''
```

```
sheet_m. write_column(chr(65 + column)
+'2',d)
book_m. close()
```

3 工程算例

3.1 算例数据选取

某混凝土双曲拱坝为一级永久建筑物,位于雅砻江大河湾干流河段,其设计坝高 305 m,为世界上已建、在建和设计中最高的双曲薄拱坝。对其坝体垂线位移监测资料进行分析,该数据已经过人工粗差去除,选取其中自 2013 年 6 月 16 日至 2015 年 9 月 28 日的测点 PL5-4 处的位移监测数据,数据每天监测 1 次。为了检验本方法的粗差识别效果,首先依据原始数据建立变形统计模型,该测点位移过程线见图 4,之后随机在其中加入 10 个粗差,见表 2,

加入粗差后的数据散点见图 5。

表 2 随机加入的粗差数据

Tab. 2 The random gross error data

日期	原始位移 测值/mm	位移测值 粗差/mm	加入粗差后位移 测值/mm
2013-07-01	0.690	-4.0	-3.310
2013-07-15	2.040	12.0	14.040
2013-08-15	3.090	-3.0	0.090
2013-10-19	6.020	-13.0	-6.980
2014-04-13	6.450	-5.0	1.450
2014-07-26	9.540	13.0	22.540
2015-02-20	11.810	-13.0	-1.190
2015-06-02	9.865	4.0	13.865
2015-08-30	12.040	13.0	25.040
2015-09-16	13.440	13.0	26.440

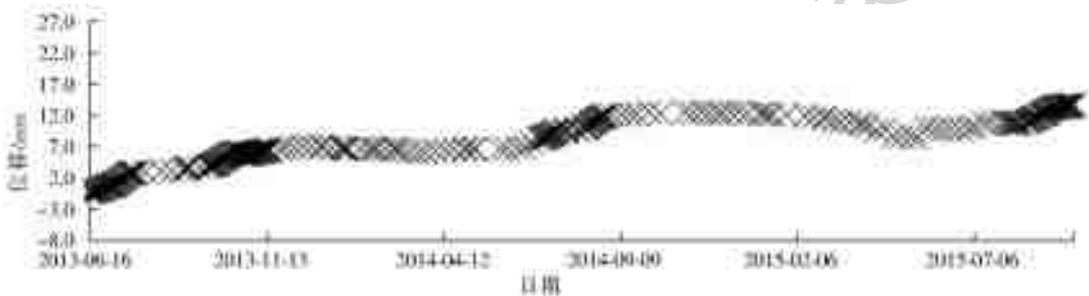


图 4 原始数据散点图

Fig. 4 The scatter diagram of original data

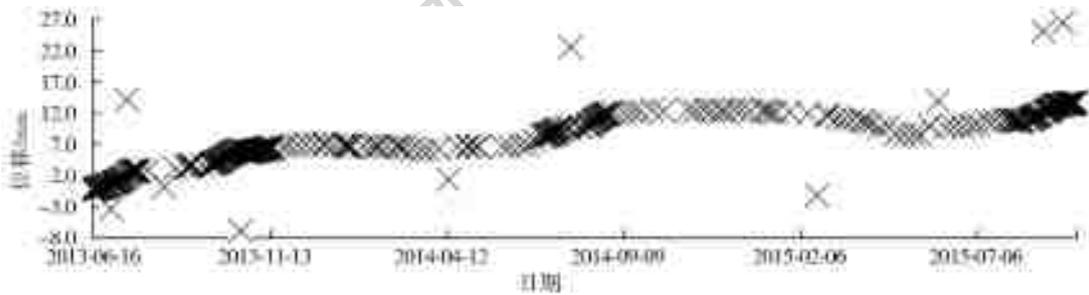


图 5 加入粗差后的数据散点图

Fig. 5 The scatter diagram of data with gross error

3.2 粗差识别方法应用与结果

将经典统计检验法中的 3σ 准则^[23]和本文粗差识别方法的识别结果进行对比,得到两种方法的处理结果见图 6、7,并根据文献[24]分别建立统计模型。

在得到统计模型的回归方程之后,可以利用复相关系数 R 和标准差 S 来衡量其有效性和精度^[25]。复相关系数 R 反映了因变量与自变量之间线性相关的密切程度与回归平方和占总离差平方和的大小, R 越大表明线性回归的效果越好,在一定程度上衡量了预报精度。此外,另一个衡量统计模型回归

精度的指标是标准差 S ,当标准差 S 越小时,其精度越高。将原始统计模型和利用两种粗差方法识别后所建立的统计模型的复相关系数和标准差进行对比,得到结果见表 3。

表 3 统计模型复相关系数与标准差对比

Tab. 3 The comparison of complex correlation coefficient and standard deviation of statistical model

指标	原始数据 统计模型	3σ 准则粗差识 别后统计模型	本文方法粗差识 别后统计模型
R	0.999	0.956	0.999
S	0.197	1.190	0.192

对加入粗差后的数据散点图进行高斯模糊和图像二值化处理,得到处理结果见图 8。

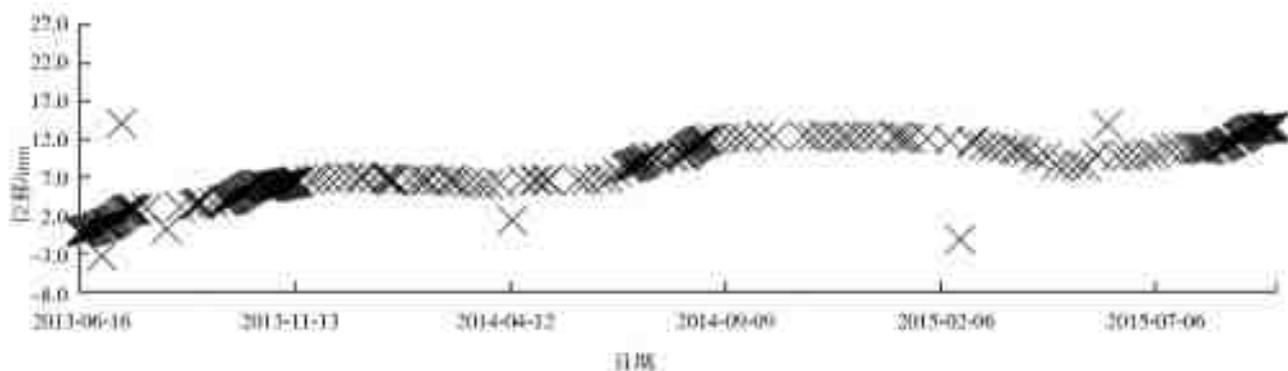


图 6 3σ 准则粗差识别后数据散点图

Fig. 6 The scatter diagram of data after identification of gross error of 3σ criteria

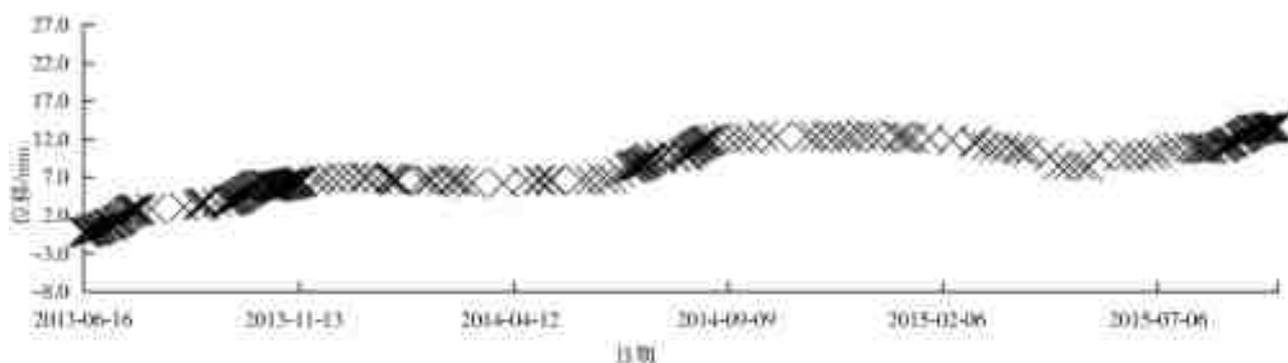


图 7 本文方法粗差识别后数据散点图

Fig. 7 The scatter diagram of data after gross error identification by the applied method of this paper



图 8 数据散点图经图像处理效果

Fig. 8 The effect of scatter diagram after image processing

由于是在原始数据基础上加入的粗差点,在运用两种方法识别粗差后得到的统计模型与依据原始数据建立的统计模型会有所偏差。依据表 3,图 4、6、7 可以看出:

(1)运用本文方法进行粗差识别,得到的统计模型回归方程的复相关系数 R 为 0.999 与依据原始数据建立的统计模型的复相关系数 R 大小相同,而利用 3σ 准则识别粗差后得到的统计模型的复相关系数 R 为 0.956,其值较小,说明回归效果较差,预报精度较低。

(2)此外,算例结果表明利用本文方法得到的统

计模型的回归方程的标准差 S 为 0.192,原始数据建立的统计模型标准差 S 值为 0.197,利用 3σ 准则识别粗差后得到的统计模型的标准差 S 为 1.190,说明利用本文方法得到的统计模型的回归方程的标准差 S 与原始数据建立的统计模型标准差 S 值更为接近,本文方法的粗差识别效果更好,其模型精度最高。

(3)由图 4、6、7 对比可以看出:本文方法比 3σ 准则对粗差的识别更加准确,能够更加彻底剔除粗差点,其粗差识别后得到的散点图与图 5 依据原始数据绘制的散点图基本一致。

4 结 论

(1)该识别方法通过模拟人类视觉,基于高斯模糊和二值化图像处理技术,相较于传统粗差识别方法,该方法识别效率更高,识别速度更快。

(2)对加入粗差的某混凝土双曲拱坝坝体 2013 年 6 月 16 日—2015 年 9 月 28 日测点 PL5-4 处的位移监测数据进行粗差识别,相较于传统 3σ 准则粗差识别方法,本文方法能够更加精准有效地识别大坝监测数据中的粗差,且识别后得出的统计模型精度更高。

本文方法仅对监测数据进行了一次识别,且本文算例中的监测数据结构较为简单,面对更加复杂多变的监测数据结构,该方法对粗差识别的效果还有待进一步验证,其算法结构也有待进一步优化。

参考文献:

- [1] 袁宏昌. 大坝变形监测自动化技术的运用与研究[J]. 农业科技与信息, 2017(1): 116-117. DOI: 10. 15979/j. cnki. cn. 62-1057/s. 2017. 01. 059.
- [2] 宋智全. 自动化技术在大坝变形监测中的应用研究[J]. 江西建材, 2018(13): 35-36. DOI: 10. 3969/j. issn. 1006-2890. 2018. 13. 017.
- [3] 苏凯. 刍议自动化技术在大坝变形监测中的应用[J]. 现代物业(中旬刊), 2019(6): 50. DOI: 10. 16141/j. cnki. 1671-8089. 2019. 06. 037.
- [4] 陶家祥,熊红阳,胡波. 论大坝安全监测数据异常值的判断方法[J]. 三峡大学学报(自然科学版), 2016, 38(6): 15-17, 41. DOI: 10. 13393/j. cnki. issn. 1672-948X. 2016. 06. 003.
- [5] 花胜强,胡少英,罗孝兵,等. 大坝位移监测中的粗差剔除方法[J]. 水力发电, 2017, 43(9): 110-112. DOI: 10. 3969/j. issn. 0559-9342. 2017. 09. 026.
- [6] 周元春,甘孝清,李端有. 大坝安全监测数据粗差识别技术研究[J]. 长江科学院院报, 2011, 28(2): 16-20. DOI: 10. 3969/j. issn. 1001-5485. 2011. 02. 003.
- [7] 魏肖怡. 大坝自动化监测数据粗差处理方法探究[J]. 科技风, 2019(31): 227. DOI: 10. 19392/j. cnki. 1671-7341. 201931202.
- [8] 程浩然,卢敏,刘存瑞,等. 基于统计学方法处理大坝安全监测中粗差问题[J]. 水利科技与经济, 2017, 23(6): 76-80. DOI: 10. 3969/j. issn. 1006-7175. 2017. 06. 017.
- [9] 陈向阳. 小波分析在 GPS 动态监测数据粗差探测中的应用[J]. 南通航运职业技术学院学报, 2013, 12(3): 47-50. DOI: 10. 3969/j. issn. 1671-9891. 2013. 03. 012.
- [10] 邓波,王毅,姜忠. 大坝变形监控数据处理的粗差识别方法及应用效果分析[J]. 水利水电技术, 2016, 47(7): 104-107. DOI: 10. 13928/j. cnki. wrahe. 2016. 07. 027.
- [11] 景继,王建. 基于抗差最小二乘的水工监测数据粗差探测[J]. 水电能源科学, 2007(6): 81-84. DOI: 10. 3969/j. issn. 1000-7709. 2007. 06. 024.
- [12] 李兴,李艳玲,张鹏,等. 基于 M 估计的改进 Pauta 准则在监测数据粗差识别中的研究及应用[J]. 中国农村水利水电, 2019(8): 133-136. DOI: 10. 3969/j. issn. 1007-2284. 2019. 08. 027.
- [13] 屈鹏. 基于 python 的地理空间数据批量处理方法研究[J]. 矿山测量, 2018, 46(1): 86-88. DOI: 10. 3969/j. issn. 1001-358X. 2018. 01. 021.
- [14] 马孝宗. 基于 Python 实现数采系统实时数据的定时采集与处理[J]. 电脑编程技巧与维护, 2019(11): 42-45. DOI: 10. 16184/j. cnki. comprg. 2019. 11. 014.
- [15] 张运玉. 基于 Python 的数据分析的研究[J]. 电脑知识与技术, 2019, 15(3): 3-4. DOI: 10. 14004/j. cnki. ckt. 2019. 3279.
- [16] 胡前防,连鹏伟,陈乾坤. Python 在统计数据处理中的应用[J]. 市场研究, 2019(8): 33-35. DOI: 10. 13999/j. cnki. scyj. 2019. 08. 009.
- [17] 韩晓冬,王浩森,王硕,等. Python 在图像处理中的应用[J]. 北京测绘, 2018, 32(3): 312-317. DOI: 10. 19580/j. cnki. 1007-3000. 2018. 03. 014.
- [18] 赵炫炫,闫晗,李保杰,等. Python Arcpy 与 Numpy 在土地集约利用中的应用:以江苏省为例[J]. 测绘地理信息, 2020, 45(1): 87-90. DOI: 10. 14188/j. 2095-6045. 2018308.
- [19] 李俊,叶松,董庆贺. Python 中函数图像快速绘制的方法[J]. 电子制作, 2014(4): 69-70. DOI: 10. 16589/j. cnki. cn11-3571/tn. 2014. 04. 236.
- [20] SAU C, Y K W, Y S HPD, et al. High dynamic range display adopting high dynamic range imaging technique[J]. IEEE Lasers and Electro-Optics Society (LEOS) Annual Meeting, 2006, 715-716. DOI: 10. 1109/LEOS. 2006. 278932.
- [21] 景天佑. 基于 Pro/ENGINEER 的 CAD 模型直接切层技术研究[D]. 西安:西安工程大学, 2012.
- [22] BERND J. Digital image processing[M]. Springer Berlin Heidelberg, 2005. DOI: 10. 1007/3-540-27563-0_1.
- [23] 何少华. 试验设计与数据处理[M]. 北京:国防科技大学出版社, 2002.
- [24] 吴中如. 水工建筑物安全监控理论及其应用[M]. 北京:高等教育出版社, 2003.
- [25] 李智录. 大坝安全监控统计模型研究[D]. 西安:西安理工大学, 2006.