

引用格式: Wang Xinrui, Jin Rui, Lin Jian, *et al.* Automatic Algorithm for Extracting Lake Boundaries in Qinghai-Tibet Plateau based on Cloudy Landsat TM/OLI Image and DEM[J]. Remote Sensing Technology and Application, 2020, 35(4): 882-892. [王鑫蕊, 晋锐, 林剑, 等. 有云 Landsat TM/OLI 影像结合 DEM 提取青藏高原湖泊边界的自动算法研究[J]. 遥感技术与应用, 2020, 35(4): 882-892.]  
doi: 10.11873/j.issn.1004-0323.2020.4.0882

# 有云 Landsat TM/OLI 影像结合 DEM 提取青藏高原湖泊边界的自动算法研究

王鑫蕊<sup>1,2</sup>, 晋锐<sup>1,3</sup>, 林剑<sup>4</sup>, 曾祥飞<sup>4</sup>, 赵泽斌<sup>1,2</sup>

(1. 中国科学院西北生态环境资源研究院, 甘肃 兰州 730000; 2. 中国科学院大学, 北京 100049;  
3. 中国科学院青藏高原地球科学卓越创新中心, 北京 100101;  
4. 湖南科技大学 知识处理与网络化制造实验室, 湖南 湘潭 411201)

**摘要:** 青藏高原的湖泊数量众多且分布广泛, 约占全国湖泊总数量和总面积的 41% 和 57%, 对于全国甚至全球的湖泊研究十分重要。遥感监测湖泊分布历来已久, 但光学遥感影像时常被云遮蔽, 难以自动化提取得到完整的湖泊边界。提出了一种有云的 Landsat TM/OLI 影像结合航天飞机雷达地形测绘数据 (Shuttle Radar Topography Mission, SRTM) 30 m 分辨率的数字高程模型 (DEM) 的湖泊完整边界的自动插值生成算法。首先, 在 Google Earth Engine 平台上, 利用 Landsat TM/OLI 影像的地表反射率 Tier1 数据, 根据其中的像元质量评价 (Pixel Quality Assessment, pixel\_qa) 属性, 结合 SRTM 30 m DEM 数据, 先剔除云、云阴影、积雪和山体区域的影响; 计算改进的归一化差异水体指数 (Modified Normalized Difference Water Index, MNDWI), 采用 Canny 边缘检测算法, 得到无云覆盖区域的已知部分湖泊边界 ( $L$ )。在本地对 DEM 进行极差滤波, 得到可能的湖泊区域; 同时, 利用 DEM 生成等高距间隔为 1 m 的等高线, 将包围着可能湖泊区域的一系列等高线自动筛选出来, 根据等高线间的包含关系建立树结构。叶子节点为最内部等高线, 记为内等高线 ( $C1$ )。由于 Landsat 和 DEM 的获取时间不同, 随着湖泊扩张或萎缩, 湖泊水位会相对于内等高线上升或下降, 对此采用不同的外等高线 ( $C2$ ) 确定方法; 随后, 建立内等高线  $C1$ 、外等高线  $C2$  与已知部分湖泊边界  $L$  之间对应点的坡度坡向关系, 插值得到未知的湖泊边界点; 最后利用最近邻法连接已知的湖泊边界点与插值得到的湖泊边界点形成完整的湖泊边界。利用相近日期的资源三号影像或无云 Landsat 影像的手工数字化湖泊边界对提取的湖泊边界进行验证, 发现两者基本重合, 且长度差百分比为 -6.81%~9.4%, 面积差百分比为 -2.11%~2.7%。表明该方法对于有云 Landsat TM/OLI 影像的湖泊边界自动化提取十分有效, 并为在 GEE 等大数据平台中自动化提取长时间序列、高时间分辨率的青藏高原湖泊边界及其时空变化分析提供了新方法。

**关键词:** 青藏高原; 湖泊边界; Landsat 影像; 有云; DEM

**中图分类号:** P941.78; TP79 **文献标志码:** A **文章编号:** 1004-0323(2020)04-0882-11

## 1 引言

湖泊是连接大气圈、生物圈、土壤圈和陆地水

圈的重要自然综合体。青藏高原幅员辽阔, 分布有数量众多且密集的内陆湖泊, 约占全国湖泊总数量和总面积的 41% 和 57%<sup>[1]</sup>, 是全国湖泊的重要组成

收稿日期: 2019-09-24; 修订日期: 2020-07-20

基金项目: 中国科学院战略性先导科技专项 (A 类) “时空三极环境” 项目 (XDA19070104), NSFC 项目 (41531174、41471357)。

作者简介: 王鑫蕊 (1995—), 女, 山西晋中人, 硕士研究生, 主要从事湖泊遥感、湖泊变化分析。E-mail: wangxinrui@lzb.ac.cn

通讯作者: 晋锐 (1979—), 女, 山西临汾人, 研究员, 主要从事水文遥感、微波遥感、数据同化及无线传感器网络研究。E-mail: jinrui@lzb.ac.cn

部分。青藏高原由于其高海拔的独特地理特征,被称为“地球第三极”<sup>[2]</sup>,人口稀少,自然环境受人类活动的影响较小,因此,对青藏高原湖泊的研究,有助于探寻自然条件下湖泊演变的规律及其与气候的相互作用,并对全球/区域水循环的研究具有重要意义。

目前,利用遥感技术监测湖泊主要集中在湖泊的面积<sup>[3-4]</sup>、水位<sup>[5-6]</sup>、水量<sup>[7]</sup>、水质<sup>[8]</sup>等方面。常见的水体提取方法有单波段阈值法<sup>[9]</sup>、谱间关系法<sup>[10]</sup>、水体指数法<sup>[11]</sup>,还有对水体的光谱、形态、水文学及空间分布等多种特征进行综合分析的决策树法<sup>[12]</sup>和基于决策树的面向对象的分析方法<sup>[13]</sup>等。其中,水体指数法是一种快速、简便、高效的水体识别方法。1996年,Mcfeeters<sup>[11]</sup>提出了归一化差异水体指数(Normalized Difference Water Index, NDWI),利用近红外和绿光波段增强了影像中的水体特征,同时消除了土壤和陆地植被特征,可用于水体的快速提取;随后,在NDWI的基础上,徐涵秋<sup>[14]</sup>提出了改进的归一化差异水体指数(Modified Normalized Difference Water Index, MNDWI),采用中红外波段代替近红外波段,取得了比NDWI更好的提取效果,尤其是在识别混合有建筑物或植被的水体特征时<sup>[15]</sup>。使用水体指数法时,阈值选取一直是个难题,阈值过小会将非水体区域误识别为水体,阈值过大则会遗漏部分水体。骆剑承等<sup>[16]</sup>提出了利用NDWI的“全局—局部”转换的迭代信息提取方法,实现了局部范围最佳阈值的确定。李均力等<sup>[17]</sup>在此方法的基础上,引入数字高程模型生成坡度信息,降低了山体阴影的干扰,建立了适合于高山地区冰川湖泊的提取算法。

湖泊作为具有一定面积的相对封闭水域,为了获取湖泊的矢量边界,常用的主要方法包括:边缘检测法<sup>[18]</sup>,即先进行边缘检测,再剔除非水域的边缘像素,最后通过连接剩余的边缘像素形成完整的水域边界;区域分割法<sup>[19]</sup>,即先进行区域分割,剔除非水体区域,再将剩余的水域边界检测出来;主动轮廓模型法<sup>[20]</sup>,即人工给出水域的种子点或初始边界,在能量函数最小的前提下,使初始边界向真实边界逼近的方法<sup>[21]</sup>;以及手工数字化方法,即通过人工数字化的方法绘制得到精确的湖泊边界,常用于验证其他方法提取湖泊边界的精度等。

在利用光学遥感影像监测湖泊面积时,往往由于云及云阴影等的遮挡,无法提取得到完整的湖泊

边界。尤其是在青藏高原,常常分布有大量的云,以2018年为例,全年覆盖青藏高原地区的Landsat-8地表反射率Tier1数据共有3 165景,其中云量在0~10%的影像共有1 002景,仅占全年总景数的31.7%,即2018年全年中青藏高原地区至少有三分之二的影像云量>10%。通过统计云量<10%的影像数量在各月份中占该月总景数的百分比,发现1~4月、10~12月的影像中约有33%~53%的影像云量<10%,而5~9月的影像中约有9%~21%的影像云量<10%。在提取湖泊时,为了避免冰和雪的影响,常常需要选择融化时期的影像,而青藏高原地区湖泊的融化时期大约在5~10月。因此,在利用湖泊融化时期的影像提取湖泊时,大部分影像都或多或少受云的影响。目前主要有3种常用去云方法:①同态滤波法:该方法可去除部分薄云<sup>[22]</sup>,然而对于较厚的云层几乎无效;②多源数据融合法:使用不同时相或不同传感器的遥感影像,用无云部分替换另一影像中被云遮挡的部分<sup>[23]</sup>,或利用SAR影像不受云的影响,基于小波变换与光学影像融合<sup>[24]</sup>,得到去云影像。该方法可以去除厚云的影响,但是可能会受限于找不到合适的无云影像或者SAR图像;③时间序列傅立叶分析法<sup>[25-26]</sup>,该方法可以将时间序列遥感数据进行简化和压缩,对时间序列遥感影像中的云及无效信息进行检测、去除和替代处理,但该方法目前主要适用于NDVI的去云处理。

所以为了利用现有的光学遥感影像、克服数据源有限的影响,结合湖泊边界位于同一高度的特点,提出了一种利用已知部分湖泊边界结合DEM数据生成完整湖泊边界的自动算法。

## 2 研究区及数据源

青藏高原那曲地区湖泊数量多,且分布密集,选择5个相邻的湖泊:达如错、切里错、懂错、江错、蓬错所在的区域作为研究区(如图1所示),用于测试该算法在一定区域内,对多个湖泊的自动化提取效果,并进行精度验证。这些湖泊面积在10~140 km<sup>2</sup>之间,且研究区内海拔范围为4 520~5 822 m。

研究区使用的光学遥感数据为来源于美国地质调查局(USGS)的一景无云2015年10月2日的、一景有云2015年9月16日的Landsat-8地表反射率Tier1数据,和一景有云1989年9月24日的、一景无云1989年10月26日的Landsat-5地表反射率Tier1

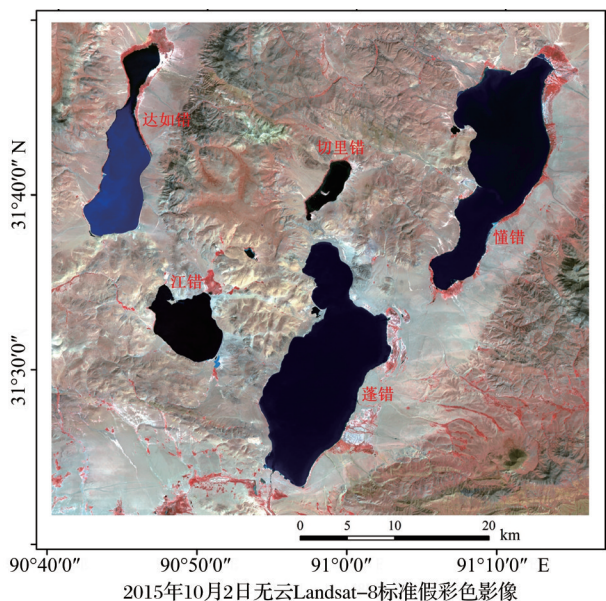


图 1 研究区

Fig.1 Research region

数据,水平分辨率均为 30 m。还使用了两景来源于中国资源卫星应用中心的资源三号(ZY-3)2015 年 9 月 28 日的无云数据,使用多光谱与全色波段的融合影像,其水平分辨率为 2.1 m。DEM 数据为来源于 USGS 的 SRTM 30 m DEM<sup>[27]</sup>,具体信息见表 1。在使用 DEM 数据之前,需先将其投影转换为对应的 Landsat 影像所处的投影坐标系。Landsat 影像和 DEM 数据主要用于湖泊边界的提取,ZY-3 影像数据用于手工数字化得到精确的湖泊边界,用来检验 Canny 边缘检测算法的效果,以及检验自动提取的最终完整湖泊边界的精度。

表 1 数据源信息

Table1 Data source information

数据种类	数据来源	空间分辨率	时间	是否有云
光学 遥感 数据	Landsat-8	30 m	2015 年 9 月 16 日	有云
			2015 年 10 月 2 日	无云
	ZY-3	2.1 m	2015 年 9 月 28 日	无云
DEM 数据	SRTM 30 m DEM	30 m	1989 年 9 月 24 日	有云
			1989 年 10 月 26 日	无云
			2000 年 2 月	不受云的 影响

### 3 湖泊边界自动提取算法

该算法主要包含 3 个步骤:①无遮蔽地区的湖泊边界提取,记为已知部分湖泊边界  $L$ ;②可能湖泊区域及其等高线树的确定;③等高线叶子节点联合已知部分湖泊边界  $L$ ,插值生成完整湖泊边界。总

体技术流程图如图 2 所示。

#### 3.1 无遮蔽地区的湖泊边界提取

##### 3.1.1 去除云、云阴影、雪和山体区域

Landsat-5 和 Landsat-8 地表反射率 Tier1 影像均包含 pixel\_qa 属性,该属性以位掩码的形式记录了由 CFMASK 算法<sup>[28]</sup>生成的像元质量评估情况。其中,pixel\_qa 属性的第 3、4、5 位分别代表云阴影、雪、云,值为 1 时代表存在云阴影、雪或云;第 6~7 位代表云的置信度,值为 0 代表置信度为 0,值为 1 代表低置信度,值为 2 代表中等置信度,值为 3 代表高置信度,置信度越高说明对云的估计结果越可靠。利用像元位值即可对云、云阴影以及积雪覆盖的区域进行掩膜处理,本研究在建立云掩膜时选择了有云且高置信度的区域。此外,由于一些山体区域可能会被误识别为水体,所以对研究区内的 DEM 数据进行坡度计算,并认为坡度  $>15^{\circ}$  的区域为山体,予以掩膜去除。

##### 3.1.2 计算改进的归一化差异水体指数

多数情况下,MNDWI 比 NDWI 的水体识别效果较好<sup>[14]</sup>,尤其是在混合有建筑物或植被的水体特征时<sup>[15]</sup>。

通过计算改进的归一化差异水体指数 MNDWI,可增大水体与非水体的区别。其公式为:

$$MNDWI = \frac{Green - MIR}{Green + MIR} \tag{1}$$

其中:Green 为绿光波段,对应 Landsat-5 的波段 2 和 Landsat-8 的波段 3;MIR 为中红外波段,对应 Landsat-5 的波段 5 和 Landsat 8 的波段 6。

##### 3.1.3 Canny 边缘检测算法

1986 年,John F Canny 提出的 Canny 边缘检测算法是一种多级边缘检测算法<sup>[29]</sup>,满足了最优边缘检测的 3 个准则:高信噪比、高定位精度、单一边缘响应;且算法相对简单,计算效率高<sup>[30-31]</sup>。其主要步骤有:高斯滤波平滑去噪;计算梯度幅值和方向;非极大值抑制;双阈值检测边缘和连接边缘。

对 MNDWI 影像进行 Canny 边缘检测即可得到较为准确的水体边界,其中被掩膜区域的边界也会被识别为水体边界,需要将其去除。接下来,判断水体边界的长度,若  $<360\text{ m}$ ,则删除该边界;若  $\geq 360\text{ m}$ ,保留其中的闭合边界,意味着保留湖泊面积至少为  $0.008\text{ 1 km}^2$  的湖泊边界;对长度  $\geq 360\text{ m}$  的不闭合边界,其中可能有部分河流和小水体的边界,以及误识别的湖泊内部水体边界,因其长度普遍较短,所以通过排除长度  $<2\text{ 000 m}$  的不闭合边



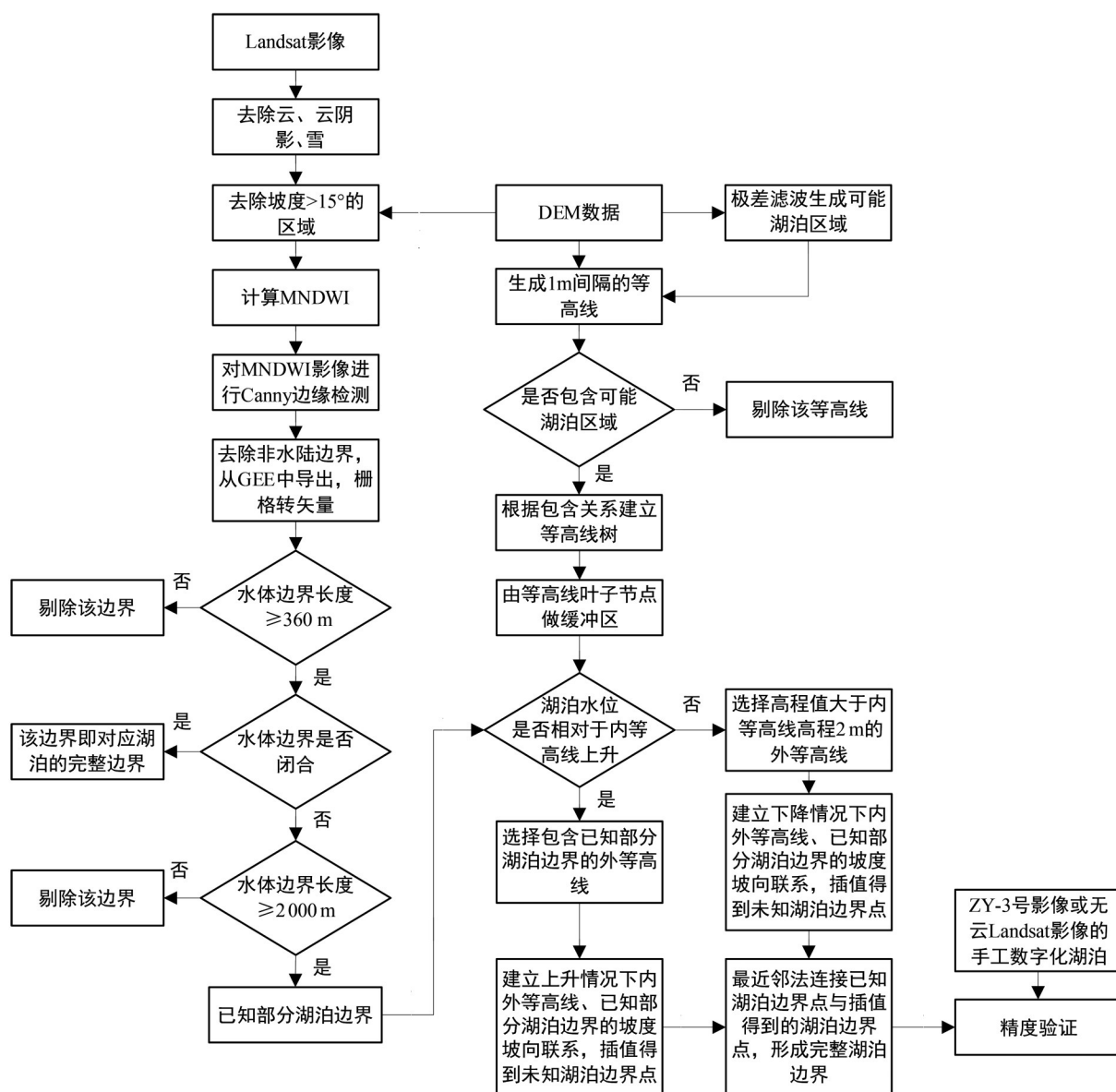


图2 技术流程图

Fig.2 Technique flow chart

界,将这些干扰性的水体边界排除掉。本算法后续将使用长度 $\geq 2\,000\text{ m}$ 的不闭合水体边界。

#### 3.1.4 边缘检测效果

选择一景无云的 Landsat-8 影像,拍摄日期为 2015 年 10 月 2 日,经过去除云、云阴影、雪和山体区域处理后,计算 MNDWI,并采用 Canny 边缘检测算法提取湖泊边界,得到长度 $\geq 360\text{ m}$ 的闭合边界与长度 $\geq 2\,000\text{ m}$ 的不闭合边界,如图 3 中绿线所示;为了验证边缘检测效果,选取一景相近日期的 ZY-3 影像(2015 年 9 月 28 日),并经过手工数字化得到湖泊边界,如图 3 中红线所示。经目视比较,可见两者十分接近,说明该方法识别湖泊边界的效果较好。

同时发现,即使是晴朗无云的影像,利用 Canny 算子进行水体边界检测时,由于湖泊边缘存在部分积雪,也导致了部分边界缺失和不连续,仍需要插值生成缺失的部分湖泊边界,并采用一定的连接算法才能获得完整湖泊边界。

### 3.2 可能湖泊区域及其等高线树的确定

#### 3.2.1 可能湖泊区域的确定

为了得到包围湖泊区域的等高线,用于湖泊边界插值,需要先确定可能的湖泊区域。由于 SRTM 30 m DEM 数据进行了空值填补处理,湖泊内部高程值一致,因此认为具有一定数目的、相同值的连通像元即为可能湖泊区域。

首先需要对 DEM 数据进行  $3\times 3$  像元的极差滤

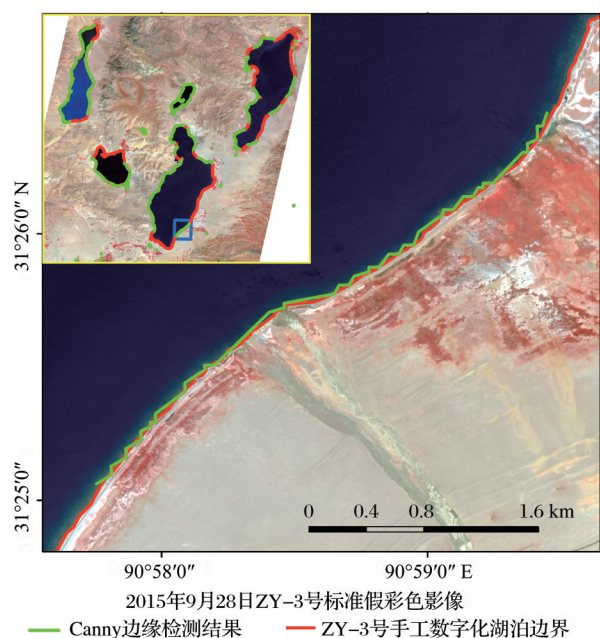


图3 Canny边缘检测算法结果

Fig.3 Edge detection results of Canny algorithm

波处理,值为0的像元即代表以该像元为中心的 $3\times 3$ 个像元值均相同;极差滤波后,选择值为0且连通像元个数 $\geq 9$ 的区域,即代表至少有25个DEM连通像元值相同,将这些像元区域转化为矢量文件,获得可能湖泊区域,其相对于实际湖泊区域向内缩小一个像元,可以被周边等高线所包围(图4)。

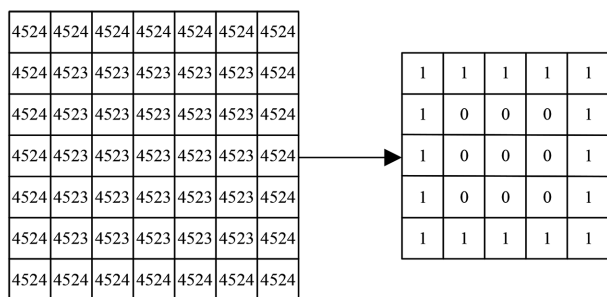


图4 DEM极差滤波示意图(单位:m)

Fig.4 Diagram of range filtering of DEM (Units: m)

### 3.2.2 等高线树的建立

由于湖泊水位多年来变化可能不大,所以为了生成更加精细的等高线,同时避免等高线过于紧凑,首先对DEM数据进行重采样,生成10 m水平间隔的DEM,进而生成1 m等高距的等高线。考虑到湖泊周围都有包围着它的闭合等高线,因此剔除不闭合等高线。然后,判断等高线是否包含可能湖泊区域中的任一要素,如不包含,则自动剔除。剩余等高线均为包含可能湖泊区域的有效等高线,每一条等高线至少包含一个可能湖泊区域。

包围同一个湖泊区域的一系列等高线互为包含关系。对于同一个湖泊,最内部等高线代表该等高线内湖泊高程值均相同,等高线越往外其高程值越大,即湖泊处于洼地的状态,也反映了湖泊周围的地势情况,等高线的形状则反映了湖泊扩张或萎缩的大致轮廓。根据包含关系,对有效等高线建立树结构,最内部等高线为叶子节点,记为C1;包含该等高线的最邻近外等高线为该等高线的父节点;以此类推,建立包含该可能湖泊区域的等高线树结构。研究区内有多个可能湖泊区域,树的任一叶子节点即为某个湖泊区域对应的最内部等高线,通过逐层寻找父节点,可找到最顶层的根节点,即包含该湖泊区域的最外部等高线;当多个湖泊区域对应同一根节点时,将建立一个等高线树;当多个湖泊区域对应不同根节点时,将建立多个等高线树。通过建立等高线树,可以方便地找到任一等高线的外等高线,便于后续插值时查找和确定合适的外等高线,记为C2。

以蓬错为例(图5),其周围密集分布的高程间隔1 m的等高线(绿线)呈相互包含关系,红线由内到外表示高程为4 529、4 539、4 549、4 559、4 567 m的等高线,4 529 m等高线即该湖泊对应等高线树的叶子节点,4 530 m等高线为4 529 m等高线的父节点,依次类推,最外层的4 567 m等高线即为该湖泊对应等高线树的根节点。图6展示了研究区内蓬错、懂错的等高线树,节点a为蓬错对应的叶子节点,节点b为懂错对应的叶子节点,节点c为蓬错和懂错共同对应的根节点。

### 3.3 完整湖泊边界的插值生成

通过无遮蔽地区的湖泊边界提取,可得到若干

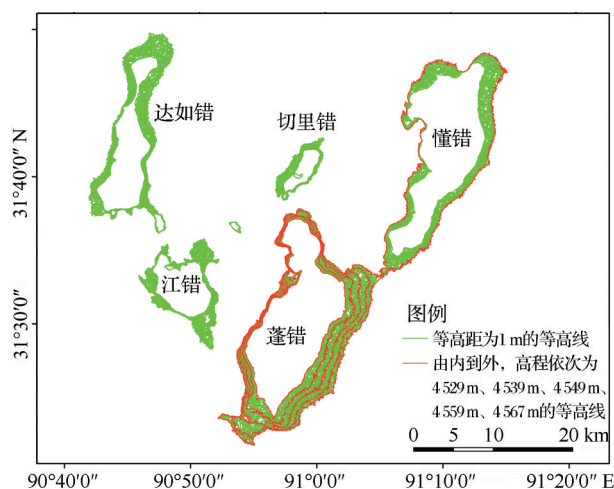


图5 等高线示意图

Fig.5 Contour Diagram

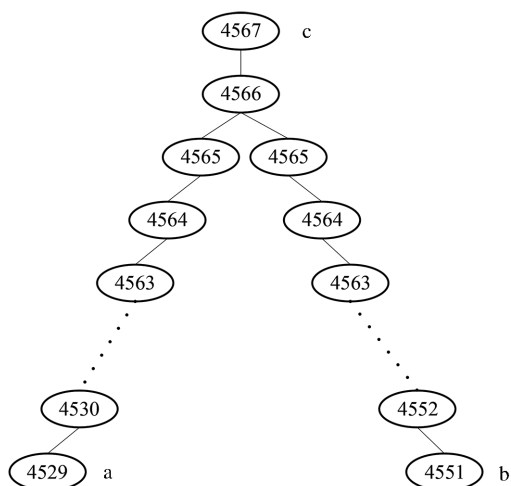


图6 部分等高线树示意图

Fig.6 Diagram of part contour tree

条长度 $\geq 2\,000\text{ m}$ 的不闭合水体边界;对于一个湖泊而言,其对应的等高线叶子节点 $C1$ 附近的不闭合水体边界即可认为是该湖泊的已知部分湖泊边界,记为 $L$ ,同时排除了一些较长的河流边界。

由于Landsat和DEM的获取时间不同,随着湖泊扩张或萎缩,湖泊水位相对于内等高线会上升或下降,即由Landsat提取的已知部分湖泊边界可能在内等高线外部或内部。为了自动判断区分这两种情况,对每个等高线叶子节点 $C1$ 做缓冲距离为自身周长5%的缓冲区,比较内外缓冲区内包含的湖泊边界线长度,若外缓冲区内边界长度 $>$ 内缓冲区内边界长度,则认为湖泊水位相对于内等高线上升;若外缓冲区内边界长度 $\leq$ 内缓冲区内边界长度,则认为湖泊水位相对于内等高线下降。针对两种情况,采用不同方法,从等高线树中选取合适的外等高线 $C2$ ,建立内外等高线与已知部分湖泊边界之间对应点的坡度坡向关系,通过插值获得未知的湖泊边界点,进而生成完整的湖泊边界。具体方法如下:

### 3.3.1 湖泊水位相对于内等高线上升

如果湖泊水位相对于内等高线 $C1$ 上升,已知部分湖泊边界 $L$ 在 $C1$ 外部,即可利用DEM计算得到 $L$ 的平均高程值。首先,选取外缓冲区内湖泊边界,边界个数 $\geq 3$ 时,进行异常边界检测,如果该边界高程平均值超过所有边界高程平均值的 $\pm 1.5$ 倍标准差范围,予以删除。在等高线树中,选择高程值比湖泊边界平均高程值至少大1 m的父等高线,由下层到上层依次判断,该外等高线是否完全包含 $L$ ,若完全包含,则将该父等高线作为湖泊插值将使

用的外等高线 $C2$ ;若不完全包含,则继续选取上层父等高线进行判断,直到最上层的父等高线;若仍不完全包含 $L$ ,则将最上层的父等高线作为湖泊插值将使用的外等高线 $C2$ 。

确定外等高线 $C2$ 后,需要建立内等高线 $C1$ 上的点到 $C2$ 上的点之间的联系,首先对 $C1$ 使用道格拉斯-普克算法<sup>[32]</sup>简化,减少 $C1$ 上节点的个数。之后,对 $C1$ 上的任一点,如点 $a$ ,做该点所处位置坡向反方向的延长线 $m$ ,寻找与 $C2$ 的交点 $b$ ; $a$ 、 $b$ 两点的高程 $H_a$ 、 $H_b$ 及距离 $L_1$ 已知,可以计算出两点之间连线的坡度。本文假设认为该条连线上的坡度值不变,同时认为 $L$ 的平均高程值为该湖泊边界的高程 $H_c$ ,若线段 $ab$ 与 $L$ 相交,则不做插值;若线段 $ab$ 与 $L$ 不相交,则进行插值,如图7所示,计算公式如下:

$$\frac{H_b - H_a}{L_1} = \frac{H_c - H_a}{L_2} \quad (2)$$

即可计算出线段 $ab$ 上湖泊边界点 $c$ 和点 $a$ 之间的距离 $L_2$ ,从而确定点 $c$ 的空间位置。

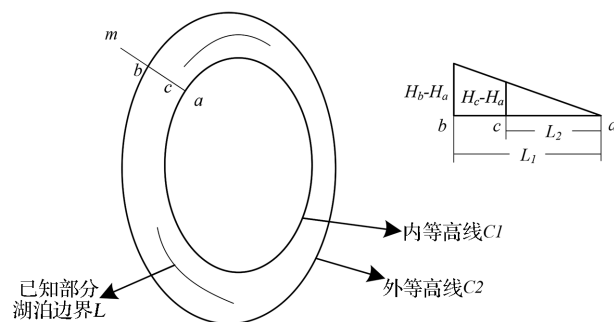


图7 湖泊水位相对于内等高线上升时,未知湖泊边界点的插值示意图

Fig.7 Interpolation of the unknown points in the lake boundary in the case of lake level rising relative to the inner contour

遍历 $C1$ 上所有点,将得到所有未知湖泊边界点的可能位置,通过最近邻法连接未知湖泊边界点和 $L$ 上的所有点。选择任一点作为起点开始连接,当剩余点数为总点数的四分之一时,在每次查找最近邻点时,需要比较当前点与下一个最近邻点的距离和当前点与起点的距离,若与起点的距离 $<$ 当前最近邻距离,则将当前点与起点相连,剩余偏差较大的点将被舍弃;否则将当前点与最近邻点相连,继续进行最近邻连接,最终得到该等高线叶子节点 $C1$ 对应的完整湖泊边界。

按照上述方法对2015年9月16日的Landsat-8影像进行完整湖泊边界的生成。如图8(a)所示,蓝色线为Canny边缘检测算法提取的长度至少为360 m



的闭合水体边界与长度至少为 2 000 m 的不闭合水体边界,红色线为插值生成的完整湖泊边界。将得到的完整湖泊边界与 ZY-3 号 2015 年 9 月 28 日影像的手工数字化湖泊边界相叠加,如图 8(b)所示,两者结果基本重合,说明本算法提取效果较好。此外,还对 5 个湖泊的边界长度和面积进行了误差统计,如表 2 所示。可见 5 个湖泊的长度差百分比介于-6.21%~

8.23% 之间,面积差百分比介于-0.06%~2.7% 之间。其中达如错、江错和蓬错的长度差较大,与手工数字化结果对比,发现误差主要来源于达如错北部、江错北部、蓬错南部等地势平坦区域(图 5)。而面积差除了达如错较大以外,其余湖泊的面积差绝对值都<0.7%,总体而言,该算法在湖泊水位相对于内等高线上升的情况下,对面积的预估较为准确。

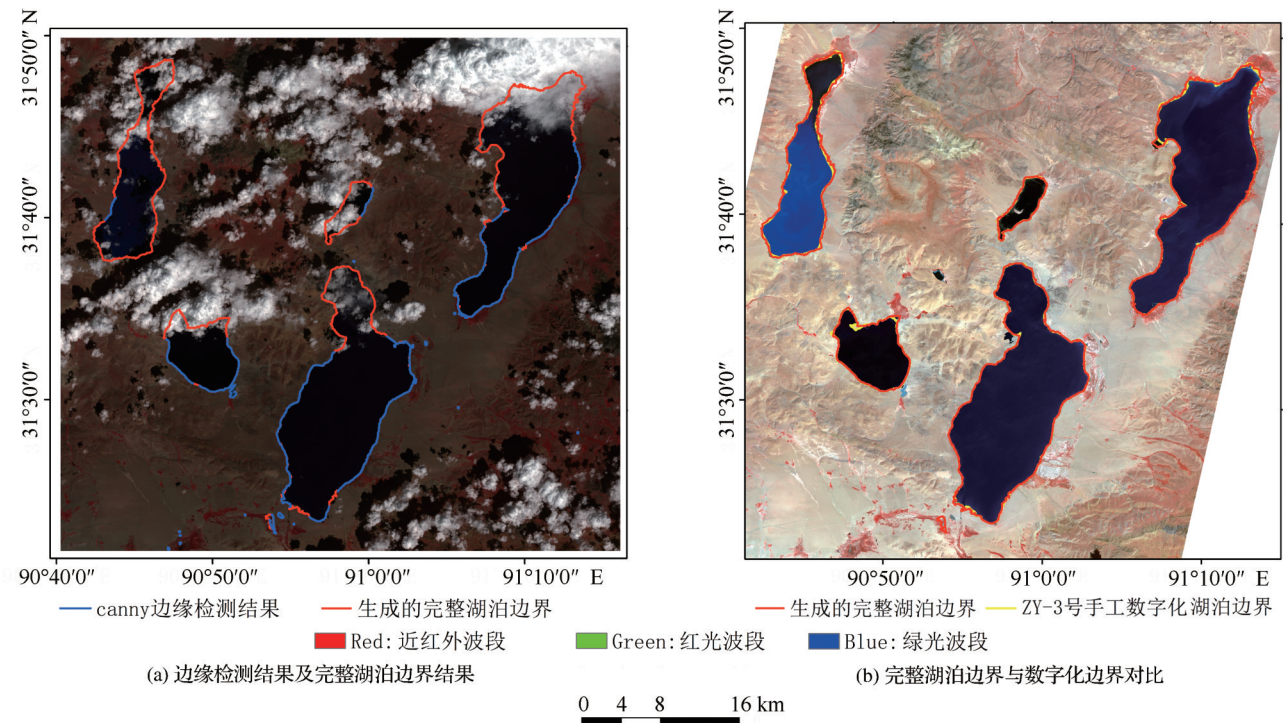


图 8 2015 年 9 月 16 日有云 Landsat-8 影像提取完整湖泊边界的验证

Fig.8 Validation of complete lake boundaries extracted from a cloudy Landsat-8 image on September 16, 2015

表 2 本算法自动提取的 2015 年湖泊边界与手工数字化边界的长度、面积差异分析

Table 2 Analysis of the differences of length and area between the automatically extracted lake boundaries by this algorithm and the manually digitized boundaries in 2015

湖泊名称	Landsat-8+DEM	ZY-3 号影像	长度差	长度差百分	Landsat-8+DEM	ZY-3 号影像	面积差	面积差
	确定长度/km	确定长度/km	/km	比/%	确定面积/km <sup>2</sup>	确定面积/km <sup>2</sup>	/km <sup>2</sup>	百分比/%
达如错	59.72	55.18	4.54	8.23	71.87	69.98	1.89	2.70
切里错	19.16	18.98	0.18	0.95	13.05	12.96	0.09	0.69
懂错	82.46	81.45	1.01	1.24	152.06	151.11	0.95	0.63
江错	29.45	31.40	-1.95	-6.21	40.65	40.43	0.22	0.54
蓬错	78.50	74.81	3.69	4.93	176.45	176.55	-0.10	-0.06

3.3.2 湖泊水位相对于内等高线下降

如已知部分湖泊边界  $L$  在内等高线  $C1$  内部,则表明湖泊实际水位相对于  $C1$  下降,无法直接根据 DEM 得到  $L$  的高程值。尝试利用  $C1$  和大于  $C1$  2 m 的外等高线  $C2$  估计  $L$  的高程,并插值生成未知湖泊边界点。若  $C1$  外层只有一条外等高线,则将该外等高线作为湖泊插值将使用的外等高线  $C2$ ;若  $C1$  外

层没有等高线则无法进行插值研究,但这种情况极少。

如图 9 所示,做  $L$  上任一点  $a$  到  $C1$  的最短距离线段,与  $C1$  交点为  $b$ ,线段  $ab$  的长度为  $L_2$ ,将线段  $ab$  向湖泊外部做延长线  $m$ ,找到  $m$  与  $C2$  的交点  $c$ ,认为线段  $ac$  上的坡度不变。根据  $b$ 、 $c$  两点的高程值  $H_b$ 、 $H_c$  及距离  $L_1$ ,可求得线段  $bc$  上的坡度,即认为该

坡度等于湖泊边界点  $a$  到  $C1$  上点  $b$  的最短距离线段的坡度。根据公式:

$$\frac{H_c - H_b}{L_1} = \frac{H_b - H_a}{L_2} \quad (3)$$

即可计算出湖泊边界点  $a$  的高程值  $H_a$ 。遍历  $L$  上每一点,计算出  $L$  上所有点的高程值,先将高程值取整,再取众数值作为该湖泊边界的高程值,记为  $H_g$ 。

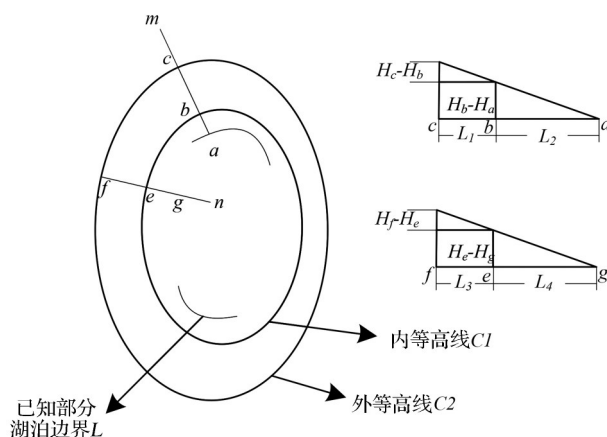


图 9 湖泊水面相对于内等高线下降时,未知湖泊边界点插值示意图

Fig.9 Interpolation of the unknown points in the lake boundary in the case of lake level falling relative to the inner contour

将  $C1$  使用道格拉斯-普克算法简化后,做  $C1$  上任一点  $e$  到  $C2$  的最短距离线段,如图 9 所示,线段与  $C2$  的交点为  $f$ ,根据  $e$ 、 $f$  两点的高程值  $H_e$ 、 $H_f$  及距离  $L_3$ ,即可求得线段  $fe$  上的坡度。做线段  $fe$  朝湖泊内部方向的一定长度的延长线  $n$ ,认为延长线  $n$  上的坡度与线段  $fe$  上的坡度一样,若延长线  $n$  与  $L$  相交,则不做插值;若延长线  $n$  与  $L$  不相交,则进行插值,湖泊边界高程  $H_g$  与内等高线高程  $H_e$  已知,根据公式:

$$\frac{H_f - H_e}{L_3} = \frac{H_e - H_g}{L_4} \quad (4)$$

即可计算出延长线  $n$  上未知湖泊边界点  $g$  距离

内等高线  $C1$  上点  $e$  的距离  $L_4$ ,从而得到湖泊边界点  $g$  的可能位置。遍历  $C1$  上每一点,将得到所有未知湖泊边界点的可能位置,通过最近邻法连接未知湖泊边界点和  $L$  上的所有点,舍弃部分偏差较大的点,最终得到该等高线叶子节点  $C1$  对应的完整湖泊边界。

按照上述方法对 1989 年 9 月 24 日的 Landsat-5 影像进行完整湖泊边界的生成。如图 10(a)所示,蓝色线为 Canny 边缘检测算法提取的长度至少为 360 m 的闭合水体边界与长度至少为 2 000 m 的不闭合水体边界,红色线为插值生成的完整湖泊边界。由于 1989 年研究区范围内没有更高分辨率的遥感影像,所以将得到的完整湖泊边界与 1989 年 10 月 26 日、无云遮挡的 Landsat-5 影像的手工数字化湖泊边界相叠加,如图 10(b)所示,两者结果基本重合,说明本算法提取效果较好。此外,还对 5 个湖泊的边界长度和面积进行了误差统计,如表 3 所示。可见 5 个湖泊的长度差百分比介于 -6.81%~9.4% 之间,面积差百分比介于 -2.11%~2.5% 之间。其中达如错、切里错、江错和蓬错的长度差较大,与手工数字化结果对比,发现误差主要来源于达如错北部、切里错西部、江错北部、蓬错南部等地势较为平缓的区域(图 5)。而面积差除了达如错、江错较大以外,其余湖泊的面积差绝对值都 <1.9%,总体而言,该算法在湖泊水位相对于内等高线下降的情况下,对面积的预估也较为准确。但是,与湖泊水位相对于内等高线上升的情况相比,该情况下的面积误差可能较大,原因应该是:该情况下无法直接根据 DEM 获得湖泊边界的高程值,而通过等高线估计得到的高程值可能不准确,导致最终结果误差较大。

## 4 结 语

该算法的提出可有效解决遥感影像中湖泊被

表 3 本算法自动提取的 1989 年湖泊边界与手工数字化边界的长度、面积差异分析

Table3 Analysis of the differences of length and area between the automatically extracted lake boundaries by this algorithm and the manually digitized boundaries in 1989

湖泊名称	Landsat-5+DEM 确定长度/km	无云影像 确定长度/km	长度差 /km	长度差 百分比/%	Landsat-5+DEM 确定面积/km <sup>2</sup>	无云影像确定 面积/km <sup>2</sup>	面积差 /km <sup>2</sup>	面积差 百分比/%
达如错	56.72	51.84	4.88	9.40	61.44	59.94	1.50	2.50
切里错	19.48	20.55	-1.07	-5.21	10.66	10.86	-0.20	-1.84
懂错	72.14	71.77	0.37	0.52	131.37	133.03	-1.66	-1.25
江错	26.68	28.63	-1.95	-6.81	36.17	36.95	-0.78	-2.11
蓬错	67.43	69.10	-1.67	-2.42	142.31	140.76	1.55	1.10



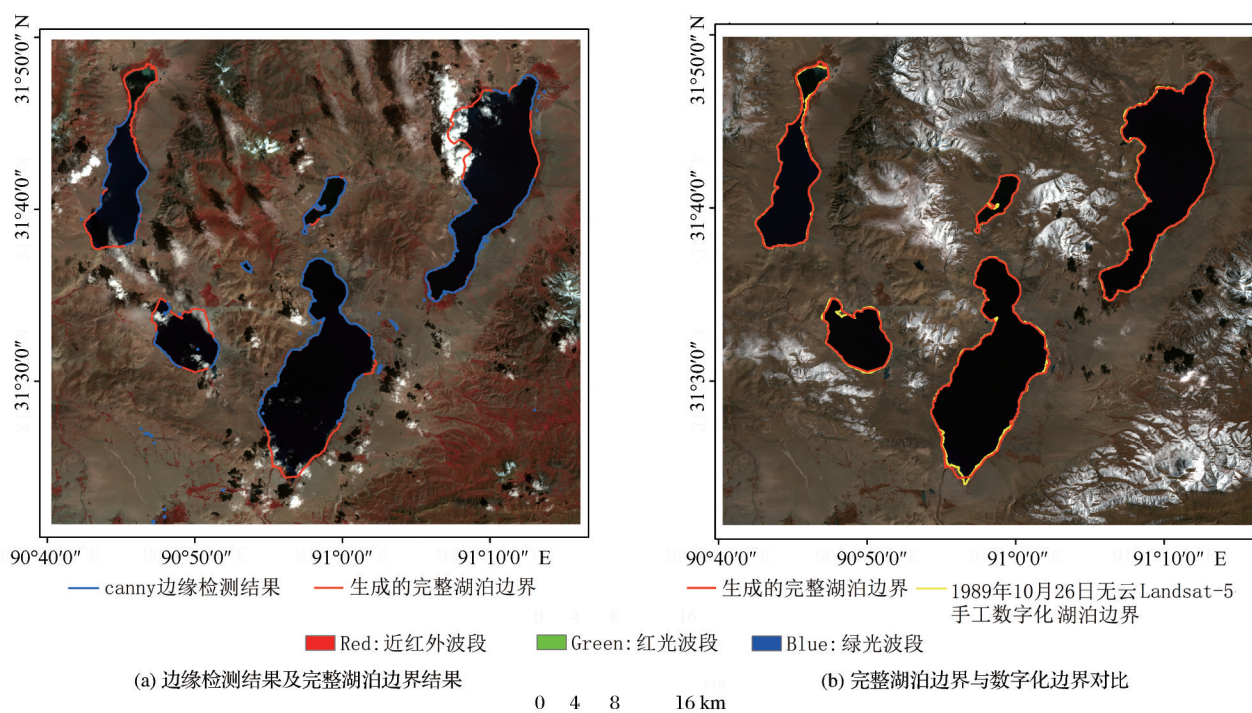


图10 1989年9月24日有云Landsat-5影像提取完整湖泊边界的验证

Fig.10 Validation of complete lake boundaries extracted from a cloudy Landsat-5 image on September 24, 1989

云部分遮挡时,如何自动提取完整湖泊边界的问题。该算法对Landsat影像去除云、云阴影、雪和山体区域后,对MNDWI计算结果进行Canny边缘检测,得到已知部分湖泊边界;进而,结合DEM数据,利用已知部分湖泊边界和生成的等高线树,根据坡度坡向不变假设,插值生成完整湖泊边界。将提取结果与手工数字化结果对比,基本重合,且精度较高。

该算法使得大量受云等因素影响的遥感影像可以得到充分利用,为研究湖泊长时间序列的年际和季节变化特征,及其对气候变化和冰冻圈变化的响应,提供了高时间分辨率的湖泊数据集;并且该算法自动化程度高,对不同时期、不同区域的影像普适性较高。

该算法的适用条件或不足之处在于:①湖泊必须具有无云遮挡的部分边界;②为了避免部分河流边界、误识别的水体边界(普遍比较短)等的干扰,本算法要求检测出的有效湖泊边界长度至少为2 000 m,才可用于完整湖泊边界的计算生成,因此只适用于面积至少为0.25 km<sup>2</sup>湖泊;③由于GEE平台上没有高精度等高线的生成算法,所以需要在本地调用arcpy包中的函数。受本地计算量的限制,该算法目前只能应用于约为3 500 km<sup>2</sup>的区域范围。未来需要在GEE平台实现高精度等高线的生成算

法;或通过提前设定湖泊感兴趣区,缩小DEM计算的空间范围等;④对于一些复杂的湖泊情况,该算法目前不一定能实现自动提取,例如湖心岛,需要边缘检测或者人工识别才能将其保留下来,但对于湖泊面积的变化分析影响不大。这些不足之处有待于未来进一步改进。

#### 参考文献(References):

- [1] Zhang G, Yao T, Chen W, *et al.* Regional Differences of Lake Evolution Across China during 1960s–2015 and Its Natural and Anthropogenic Causes[J]. *Remote sensing of Environment*, 2019, 221: 386–404.
- [2] Madsen D B. Conceptualizing the Tibetan Plateau: Environmental Constraints on the Peopling of the “Third Pole” [J]. *Archaeological Research in Asia* 5, 2016, 5: 24–32.
- [3] Ma R, Yang G, Duan H, *et al.* China’s Lakes at Present: Number, Area and Spatial Distribution [J]. *Science China Earth Sciences*, 2011, 54(2): 283–289.[马荣华, 杨桂山, 段洪涛, 等. 中国湖泊的数量、面积与空间分布[J]. *中国科学: 地球科学*, 2011, 41(3): 394–401.]
- [4] Zhang G, Luo W, Chen W, *et al.* A Fobust but Variable Lake Expansion on the Tibetan Plateau[J]. *Science Bulletin*, 2019, 64(18): 1306–1309.
- [5] Crétau J F, Jelinski W, Calmant S, *et al.* SOLS: A Lake Database to Monitor in the Near Real Time Water Level and Storage Variations from Remote Sensing Data[J]. *Advances in Space Research*, 2011, 47(9): 1497–1507.
- [6] Zhang G, Xie H, Kang S, *et al.* Monitoring Lake Level Changes on the Tibetan Plateau Using ICESat Altimetry Data

- (2003~2009) [J]. Remote Sensing of Environment, 2011, 115(7): 1733-1742.
- [7] Zhang G, Yao T, Shum C K, *et al.* Lake Volume and Groundwater Storage Variations in Tibetan Plateau's Endorheic basin: Water Mass Balance in the TP[J]. Geophysical Research Letters, 2017, 44(11): 5550-5560.
- [8] Koponen S, Pulliainen J, Kallio K, *et al.* Lake Water Quality Classification with Airborne Hyperspectral Spectrometer and Simulated MERIS Data[J]. Remote Sensing of Environment, 2002, 79(1): 51-59.
- [9] Jupp D L B, Mayo K K, Kucher D A, *et al.* Landsat based interpretation of the Cairns section of the Great Barrier Reef Marine Park[M]. Canberra, ACT, CSIRO Division of Water & Land Resources, 1985.
- [10] Zhou C H, Du Y Y, Luo J C. A Description Model based on Knowledge for Automatically Recognizing Water from NOAA/AVHRR[J]. Journal of Natural Disasters, 1996, 5(3): 100-108.[周成虎, 杜艳艳, 骆剑承. 基于知识的 AVHRR 影像的水体自动识别方法与模型研究[J]. 自然灾害学报, 1996, 5(3): 100-108.]
- [11] Mcfeeters S K. The Use of the Normalized Difference Water Index(NDWI) in the Delineation of Open Water Features[J]. International Journal of Remote Sensing, 1996, 17(7): 1425-1432.
- [12] Du Jinkang, Huang Yongsheng, Feng Xuezi, *et al.* Study on Water Bodies Extraction and Classification from SPOT Image [J]. Journal of Remote Sensing, 2001, 5(3): 214-219.[都金康, 黄永胜, 冯学智, 等. SPOT 卫星影像的水体提取方法及分类研究[J]. 遥感学报, 2001, 5(3): 214-219.]
- [13] Shen Jinxiang, Yang Liao, Chen Xi, *et al.* A Method for Object-oriented Automatic Extraction of Lakes in the Mountain Area from Remote Sensing Image [J]. Remote Sensing for Land & Resources, 2012, 24(3): 84-91.[沈金祥, 杨辽, 陈曦, 等. 面向对象的山区湖泊信息自动提取方法[J]. 国土资源遥感, 2012, 24(3): 84-91.]
- [14] Xu H Q. Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery[J]. International Journal of Remote Sensing, 2006, 27(14): 3025-3033.
- [15] Singh K V, Setia R, Sahoo S, *et al.* Evaluation of NDWI and MNDWI for Assessment of Waterlogging by Integrating Digital Elevation Model and Groundwater Level[J]. Geocarto International, 2015, 30(6): 650-661.
- [16] Luo Jianchen, Sheng Yongwei, Shen Zhanfeng, *et al.* Automatic and High-precise Extraction for Water Information from Multispectral Images with the Step-by-Step Iterative Transformation Mechanism [J]. Journal of Remote Sensing, 2009, 13(4): 604-615.[骆剑承, 盛永伟, 沈占锋, 等. 分步迭代的多光谱遥感水体信息高精度自动提取[J]. 遥感学报, 2009, 13(4): 604-615]
- [17] Li Junli, Sheng Yongwei, Luo Jiancheng. Automatic Extraction of Himalayan Glacial Lakes with Remote Sensing [J]. Journal of Remote Sensing, 2011, 15(1): 29-43.[李均力, 盛永伟, 骆剑承. 喜马拉雅山区冰湖信息的遥感自动化提取[J]. 遥感学报, 2011, 15(1): 29-43.]
- [18] Liu H, Jezek K. Automated Extraction of Coastline from Satellite Imagery by Integrating Canny Edge Detection and Locally Adaptive Thresholding Methods[J]. International journal of remote sensing, 2004, 25(5): 937-958.
- [19] Setiawan B D, Rusydi A N, Pradityo K. Lake edge detection using Canny algorithm and Otsu thresholding[C]// 2017 International Symposium on Geoinformatics (ISyG), IEEE, 2017: 72-76, doi: 10.1109/ISYG.2017.8280676.
- [20] Kumar R. Snakes: Active Contour Models [J]. International Journal of Computer Vision, 1988, 1(4): 321-331.
- [21] Zhu Shulong, Meng Weican, Zhu Baoshan. Irregular Water Boundary Extraction Using GVF Snake [J]. Journal of Remote Sensing, 2013, 17(4): 742-758.[朱述龙, 孟伟灿, 朱宝山. 运用 GVF Snake 算法提取水域的不规则边界[J]. 遥感学报, 2013, 17(4): 742-758.]
- [22] Shen H, Li H, Yan Q, *et al.* An Effective Thin Cloud Removal Procedure for Visible Remote Sensing Images [J]. Isprs Journal of Photogrammetry & Remote Sensing, 2014, 96(11): 224-235.
- [23] Wang Hui, Tan Bing, Shen Zhiyun. The Processing Technology of Removing Clouds Image based on the Multi-resource RS Image [J]. Journal of the PLA Institute of Surveying & Mapping, 2001, 18(3): 195-198.[王惠, 谭兵, 沈志云. 多源遥感影像的去云层处理[J]. 测绘学院学报, 2001, 18(3): 195-198.]
- [24] Eckardt R, Berger C, Thiel C, *et al.* Removal of Optically Thick Clouds from Multi-spectral Satellite Images Using Multi-frequency SAR Data [J]. Remote Sensing, 2013, 5(6): 2973-3006.
- [25] Roerink G J, Menenti M, Verhoef W. Reconstructing Cloud-free NDVI Composites Using Fourier Analysis of Time Series [J]. International Journal of Remote Sensing, 2000, 21(9): 1911-1917.
- [26] Wang Dan, Jiang Xiaoguang, Tang Lingli, *et al.* The Application of Time-series Fourier Analysis to Reconstructing Cloud-free NDVI Images [J]. Remote Sensing for Land & Resources, 2005, 17(2): 29-32.[王丹, 姜小光, 唐伶俐, 等. 利用时间序列傅立叶分析重构无云 NDVI 图像[J]. 国土资源遥感, 2005, 17(2): 29-32.]
- [27] Farr T G, Rosen P A, Caro E, *et al.* The Shuttle Radar Topography Mission [J]. Reviews of Geophysics, 2007, 45(2): RG2004, doi:10.1029/2005RG000183.
- [28] Foga S, Scaramuzza P L, Guo S, *et al.* Cloud Detection Algorithm Comparison and Validation for Operational Landsat Data Products [J]. Remote Sensing of Environment, 2017, 194: 379-390.
- [29] Canny J. A Computation Approach to Edge Detection [J]. IEEE Trans Pattern Anal Mach Intell, 1986, 8(6): 670-700.
- [30] Ding L, Goshtasby A. On the Canny Edge Detector [J]. Pattern Recognition, 2001, 34(3): 721-725.
- [31] Mcilhagga W. The Canny Edge Detector Revisited [J]. International Journal of Computer Vision, 2011, 91(3): 251-261.
- [32] Douglas D H, Peucker T K. Algorithms for the Reduction of the Number of Points Required to Represent a Digitized Line or Its Caricature [J]. Cartographica: the International Journal for Geographic Information and Geovisualization, 1973, 10(2): 112-122.

## Automatic Algorithm for Extracting Lake Boundaries in Qinghai-Tibet Plateau based on Cloudy Landsat TM/OLI Image and DEM

Wang Xinrui<sup>1,2</sup>, Jin Rui<sup>1,3</sup>, Lin Jian<sup>4</sup>, Zeng Xiangfei<sup>4</sup>, Zhao Zebin<sup>1,2</sup>

(1. *Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China;*

*2. University of Chinese Academy of Sciences, Beijing 100049, China;*

*3. Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing 100101, China;*

*4. Key Laboratory Knowledge Processing and Networked Manufacturing, Hunan University of Science and Technology, Xiangtan 411201, China)*

**Abstract:** Lakes in the Qinghai-Tibet Plateau are numerous and widely distributed, accounting for 41% and 57% of the total number and area of lakes in China, which are very important for the study of lakes in the whole country and even in the whole world. Remote sensing has been used to monitor the lake distribution for a long time, but optical remote sensing images are often obscured by clouds, from which it's impossible to automatically extract complete lake boundaries. An automatic interpolation algorithm for lake boundary generation based on cloudy Landsat TM/OLI image and Shuttle Radar Topography Mission (SRTM) 30 m resolution Digital Elevation Model (DEM) is proposed. Firstly, supported by the platform of Google Earth Engine, the tier1 data of Landsat TM/OLI images are used to eliminate the effects of cloud, cloud shadow, snow and mountain area, based on the Pixel Quality Assessment (pixel\_qa) attribute and SRTM 30 m DEM. Then, the Modified Normalized Difference Water Index (MNDWI) is calculated, and the Canny edge detection algorithm are used to obtain the known part of the lake boundary ( $L$ ) in cloud-free areas. The possible lake areas are obtained by range filtering of DEM locally. At the same time, DEM is used to generate contours with an isometric interval of 1 m, and a series of contours surrounding the possible lake area are automatically screened out. The tree structure is established according to the inclusion relationship between contours. The leaf nodes are the innermost contours, which are recorded as inner contours ( $C1$ ). Because the acquisition time of Landsat and DEM is different, with the lake expanding or shrinking, the lake water surface will rise or fall relative to the inner contour. Different methods of determining the outer contour ( $C2$ ) are adopted. Subsequently, the slope-aspect relationship between the inner contour  $C1$  and the outer contour  $C2$  and the known part of the lake boundary  $L$  is established, and the unknown lake boundary points are interpolated. Finally, the nearest neighbor method is used to connect the known lake boundary points with the interpolated Lake boundary points to form a complete lake boundary. The extracted lake boundaries were validated by visual digitized lake boundaries from ZiYuan-3 image or cloud-free Landsat image on the near date. It is found that they are basically coincided, and the percentage of differences in length and area are  $-6.81\% \sim 9.4\%$  and  $-2.11\% \sim 2.7\%$  respectively. It shows that this method is very effective for automatic extraction of Lake boundary from cloudy Landsat TM/OLI images, and provides a new method for automatic extraction of long time series Lake boundary and its temporal and spatial variation analysis in the Qinghai-Tibet Plateau on GEE and other big data platforms.

**Key words:** Qinghai-Tibet plateau; Lake boundary; Landsat; Cloud; DEM