

## 基于高分三号卫星数据的水体自动提取及应用\*

崔倩<sup>1</sup>, 毛旭东<sup>2</sup>, 陈德清<sup>1</sup>, 瞿孟<sup>2</sup>, 白珏莹<sup>2</sup>

(1. 水利部信息中心, 北京 100053; 2. 武汉大学, 湖北武汉 430072)

**摘要:**【目的】高分三号(GF-3)卫星具有全天时、全天候对地观测能力, 基于高分三号雷达数据开展水体自动提取及业务运行研究, 是洪涝监测、旱情监测等水资源管理的技术支撑。【方法】文章提出了一种基于阈值自动分割的水体自动化提取算法, 通过影像直方图分布特征自动确定分割阈值, 自动化提取水体。针对全国大范围业务监测需求, 设计提出数据订阅接收、数据预处理、水体提取、后处理和产品入库、数据分析和产品展示的全流程业务运行系统。【结果】以云南省阳宗海、江苏省洪泽湖、北京市密云水库等3个典型水体验证算法的有效性, 比较了不同极化数据水体提取结果。(1)直方图呈现典型的双峰特征分布, 水体和背景像元被自动化寻找的阈值有效分割;(2)水体提取的精度对极化方式的选取并不敏感。【结论】采用GF-3卫星数据进行水体自动提取, 能有效监测水资源, 为洪涝、旱情等农田水利管理提供数据支持。

**关键词:** 高分三号; 水体提取; 自动化阈值; 业务化

DOI: 10.12105/j.issn.1672-0423.20190506

## 0 引言

我国洪涝和干旱灾害和水资源短缺等水问题突出, 地面监测系统目前还不能实现全覆盖。高分三号卫星具有全天候全天时对地观测能力, 可以不分昼夜、阴雨天气进行对地观测。针对洪涝灾害监测、水资源管理等水利应用需求, 基于可以常态化接收的国产雷达数据, 实现水体的自动化提取, 可以补充地面监测能力的不足, 在提升我国水利监测能力方面发挥了重要作用。

常用的雷达水体提取技术包括目视解译<sup>[1]</sup>、监督分类<sup>[2-4]</sup>、纹理特征分析<sup>[4-5]</sup>、动态轮廓模型<sup>[6-7]</sup>、阈值分割等。监督分类法通过构造灰度、纹理等特征, 标注样本训练支持向量机等监督分类器提取目标地物, 算法稳定有效, 但需要人工标注样本, 不适用于大批量水体提取工作。纹理特征分析方法从地物粗糙度的差异入手, 利用灰度共生矩阵等纹理特征提取地物, 提取精度高, 但计算量大, 并且效果对纹理窗口尺寸依赖较大。动态轮廓模型将分割图像提取水体的问题转换为求解能量泛函最小值问题, 对噪声有很强的鲁棒性, 但对参数设置敏感且效率欠佳。阈值分割法是应用最为广泛的水体提取方

收稿日期: 2019-09-29

第一作者简介: 崔倩(1986—), 博士、高级工程师。研究方向: 水文遥感。Email: cuiqian@mwr.gov.cn

\* 基金项目: 国家重点研发计划“国家水资源立体监测体系与遥感技术应用”(2017YFC0405800)

法，其原理设定阈值分割雷达影像，将图像中小于阈值部分标记为水体，大于阈值部分标记为背景，获得水体范围二值图。该方法易于实现，速度快，但难点在于阈值的确定。最直接的方式通过人机交互的方式选取阈值，但当处理大批量数据时效率低下，并且获取的结果主观性强，可追踪性差。

针对人机交互阈值选取方法上的不便，自动化阈值选取方法被应用到水体提取算法中。常用的自动阈值分割选取方法有双峰法、OTSU 方法<sup>[8-9]</sup>、最大熵阈值法<sup>[10]</sup>和最小误差阈值法。

出于自动化、业务化的考虑，所应用的雷达水体提取方法应具有简单、稳定、快速的特点，并具有满足业务需求的提取精度。结合上述常用水体提取方法，文章采用双峰法进行业务化水体提取工作。直观上说，双峰法寻找阈值的过程，是根据雷达影像后向散射系数统计分布直方图呈现出的双峰值特征、在两峰之间寻找波谷的过程。

值得注意的是，阈值自动选取方法理论上的可行性依赖水体与背景地物的反射特征差异，即雷达影像中水体像元与背景像元后向散射系数大小的显著差异，反应在统计分布直方图上即图形的双峰形走势。然而，当影像中水体地物所占比例很小时，在直方图中水体像元的特征被背景像元掩盖、图形不再呈现出双峰形，水体提取效果会受到很大影响。为了避免这一情况的出现，先使用全国水利普查成果湖库水体缓冲区矢量对影像做裁剪，再在裁剪结果上用双峰法提取水体，保证业务化雷达水体提取方法的精度。

## 1 雷达水体自动化提取算法

### 1.1 后向散射系数分布模型

与可见光和近红外主要对地物的化学和热性质敏感不同，微波遥感主要对地物的表面、亚表面及覆盖物的物理性质和介电性质敏感。含水量和粗糙度是决定雷达后向散射的重要因素，一般含水量越大，粗糙度越大，后向散射越强。另外，不同频率、入射角和不同极化方式的后向散射也有明显不同。

对于水体来说，未扰动水面的后向散射系数非常低，在所有雷达波段都接近于镜面反射，微波能量对于高介电常数几乎以等于入射角的方向反射出去。就粗糙度看，平静水面相对于雷达波来说几乎是水平的，因此水的雷达信号非常弱，区别于其他地物，在影像中呈现为低后向散射系数的均匀区域。在水体提取任务中，通常只对某个像元是否属于水体感兴趣，对于非水体的背景像元的具体地物类型并不在意。考虑到水体像元的低后向散射特性，将像元划分为水体与背景 2 类，并进一步讨论这 2 类像元的特征。

用后向散射系数 ( $\sigma_0$ ) 描述 SAR 影像的散射特征，则整个研究区域所有像元的  $\sigma_0$  组成的样本集合可以视作水体与背景两部分的非交并集<sup>[11]</sup>。记水体像元的后向散射系数分布概率密度函数为  $p(\sigma_0|W)$ ，背景像元的后向散射系数分布概率密度函数为  $p(\sigma_0|\bar{W})$ ，研究区域后向散射系数边际分布的概率密度函数为  $p(\sigma_0)$ ，则以上表述可以概括为：

$$p(\sigma_0) = p(\sigma_0|W)p(W) + p(\sigma_0|\bar{W})p(\bar{W}) \quad (1)$$

2019年10月

式(1)中,  $p(W)$  与  $p(\bar{W})$  分别表示水体与背景像元所占比例,  $p(W) + p(\bar{W}) = 1$ 。即  $p(\sigma_0)$  可视为2类像元的混合分布。该分布形式与影像直方图的双峰形走势是一致的。

## 1.2 自动化阈值寻找算法

统计学上认为直方图是对概率密度函数的估计<sup>[12]</sup>。经过预处理后的研究对象, 逐个统计单一研究区域内像元值的最小值和最大值, 在值域内定义相应对象的后向散射系数直方图。直方图计算需要的参数为其每个条带的宽度, 称为带宽。带宽的计算公式为<sup>[12]</sup>:

$$bandwidth = \frac{2.6IQR}{n^{\frac{1}{3}}} \quad (2)$$

式(2)中,  $n$  为统计样本个数,  $IQR$  为样本的四分位差。研究表明, 由式(2)给出的带宽适用于 Gauss 分布样本数据。

对具有明显峰谷形状的图像直方图来说, 阈值法的分割效果较好。如果图像的直方图呈现出多峰多谷形状, 则存在多个直方图分割阈值, 此时图像可分割成多个目标区域。在后向散射系数直方图上紧邻目标或背景峰值所发生畸变的地方可以认为是最佳分割阈值。考虑到水体提取是在全国水利普查成果矢量缓冲区范围内进行, 缓冲区范围内地物类型较少, 后向散射系数直方图分布较为简单, 表现为典型的双峰状。水体像元由于其低散射特性主要分布在直方图的左侧, 因此只需要确定一个合适的阈值分割目标水体与背景, 而不用考虑背景中不同地物之间的分割阈值。

以直方图横坐标第1个灰度级为初始峰值, 逐步比较每个灰度级所对应的像元个数, 通过循环迭代遍历整个直方图, 寻找直方图上最大峰值。利用公式(3)将直方图在灰度级 HistSize 范围内平均划分为左、中、右3部分, 公式(4)定义了遍历步长。

$$M = (df_{\max} - df_{\min}) / 3 \quad (3)$$

$$d_{\text{step}} = (df_{\max} - df_{\min}) / (\text{HistSize} - 1.0) \quad (4)$$

式(3)~(4)中,  $df_{\max}$  表示研究区域内像元灰度级的最大值,  $df_{\min}$  表示研究区域内像元灰度级的最小值。

在后向散射系数直方图上, 目标或背景峰值位置分布有以下3种情况, 针对不同情况, 有相应的分割阈值判断方法。

(1) 最大的峰值位置位于直方图的右侧(直方图曲线最大峰值位置  $N_{\max\text{Index}} > 2\text{HistSize}/3$ ), 说明研究区域内目标水体区域面积较大, 此时以  $d_{\text{step}}$  为步长, 从峰值处往左侧遍历, 把直方图曲线上第1次发生突变点判断为分割阈值。

(2) 最大的峰值位置位于直方图的左侧(直方图曲线最大峰值位置  $N_{\max\text{Index}} \leq \text{HistSize}/3$ ), 则说明研究区域内背景区域面积较大, 此时以  $d_{\text{step}}$  为步长, 从峰值处往右侧遍历找出分割阈值。考虑到右侧区域内可能还存在其他背景地物的影响, 把距该峰值最远处的突变点判断为分割阈值。

(3) 最大的峰值位置位于直方图的中部(直方图曲线最大峰值位置  $\text{HistSize}/3 < N_{\max\text{Index}} \leq 2\text{HistSize}/3$ ), 如果该峰值右侧仍然存在一个较明显的峰值, 则该最大峰值为背景峰值, 此时采用步骤(2)中的方法寻找分割阈值, 否则认为该最大峰值为目标峰值, 此时采用

步骤 (1) 中的方法寻找分割阈值。

## 2 实验与分析

### 2.1 实验区域与数据

为了验证雷达水体提取算法的有效性,在全国范围内选取了 3 个典型湖库进行实验,分别是云南省阳宗海、江苏省洪泽湖和北京市密云水库。综合考虑影像的标称分辨率、成像带宽、湖库覆盖程度以及数据处理难度,精细条带、标准条带模式下的中高分辨率、成像带宽为 50~150 km 的影像数据更适用于业务化水体提取需求。该文选取了 3 座湖库的 GF-3 精细条带 2 影像用于实验,实验数据如表 1 所示。

表 1 GF-3 影像数据  
Table 1 GF-3 images data

序号	影像名称
1	GF3_KAS_FSIL_006639_E103.2_N25.2_20171113_L1A_HHHV_L10002764936
2	GF3_KAS_FSIL_006105_E117.0_N40.3_20171007_L1A_HHHV_L10002658489
3	GF3_KAS_FSIL_006105_E118.6_N33.4_20171007_L1A_HHHV_L10002658482
4	GF3_MYN_QPSI_002162_E103.0_N24.9_20170106_L1A_AHV_L10002101437

### 2.2 实验区域与数据

图 1~3 分别为阳宗海、洪泽湖和密云水库的算法处理结果。图 1a、图 2a 和图 3a 为裁剪后的后向散射系数图,使用的是精细条带 2 模式 HH 单极化影像。可以看出与非水地物相比,水体因其低值而呈现为暗色区域。图 1b、图 2b 和图 3b 为水体提取结果图,从影像直观来看,水体范围得到了有效提取。湖泊、水库地区多云雾天气,尤其是云贵川等地,常年有云,光学影像难以做到常态化监测,对比之下利用雷达影像提取水体更具优势。此外,水体提取结果图中存在若干噪点,可以通过后续的形态学处理消除。图 1c、图 2c 和图 3c 为后向散射系数的统计分布直方图,经过裁剪后的影像与原始影像相比,区域内水体与背景像元的比例更加均衡,因此直方图呈现典型的双峰形走势,为自动阈值寻找算法的有效性提供了保障,其中红色直线表示自动阈值寻找算法所寻阈值。从图上看,该方法在 3 幅影像中均较为准确地找到了直方图波谷所在的位置,有效分割了影像中的水体与背景像元,证明了自动化阈值寻找算法的有效性。

通过寻找直方图波谷确定最佳分割阈值的方法的合理性在于直方图的条件概率密度可正确模拟类的分布的假设<sup>[13]</sup>。为了从理论上验证雷达水体提取算法的有效性,对后向散射系数的统计分布模型做进一步的估计。假设影像中水体与背景像元均服从高斯分布,结合公式(1),使用 Levenberg-Marquardt 方法集合直方图估计模型参数得到影像后向散射系数的理论分布<sup>[14]</sup>,叠加在直方图上,显示为图中绿色的曲线。如果该曲线反映了后向散射系数的真实统计规律,那么理论上最佳的分割阈值应处在曲线中 2 个高斯分布众数之间的波谷处。从图上看,该算法求得阈值所在的红线与理论分布绿色曲线的波谷位置相近,相距均不超过 2 dB,说明了阈值选取的合理性。

2019年10月

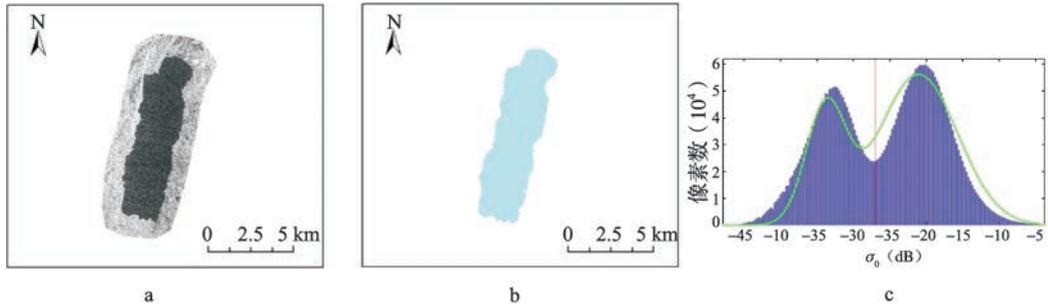


图1 阳宗海实验区: a 后向散射系数影像; b 水体提取结果; c 后向散射系数直方图  
Fig.1 Yangzonghai Lake area

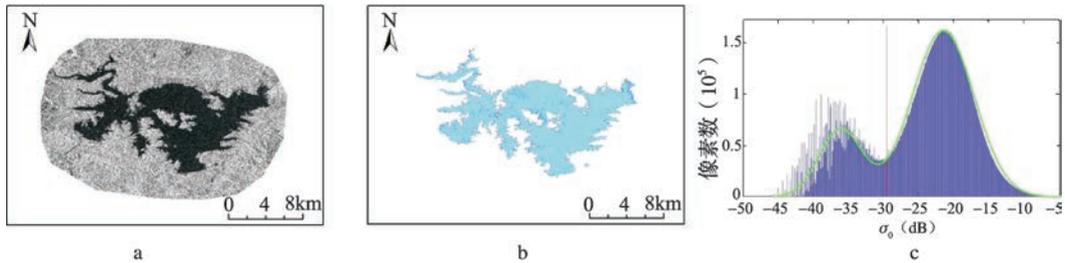


图2 洪泽湖实验区: a 后向散射系数影像; b 水体提取结果; c 后向散射系数直方图  
Fig.2 Hongze Lake area

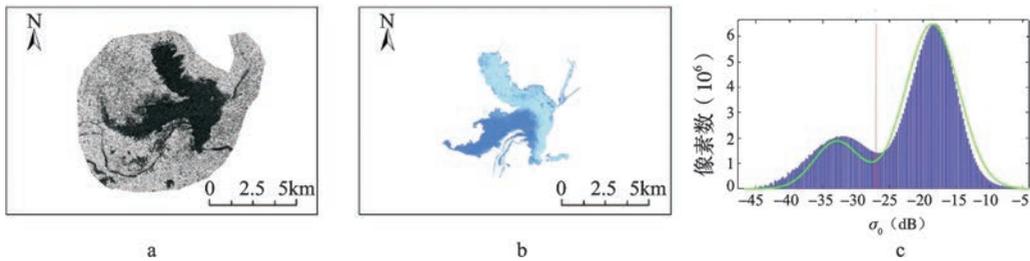


图3 密云水库实验区: a 后向散射系数影像; b 水体提取结果; c 后向散射系数直方图  
Fig.3 Miyun Reservoir area

### 2.3 极化方式的选取

在实验中用于业务化水体提取场景的精细条带、标准条带模式数据均是多极化方式。该文将进一步分析极化方式对水体提取结果的影响。

为了减少实验的不确定性,选用阳宗海区域的全极化条带模式数据,分别尝试用4种极化影像进行水体提取,实验结果如图4。从图4看,除HH极化影像因成像质量问题噪声较大导致提取效果不佳外,其余3种极化影像均有效地提取出了水体区域,其差异部分主要集中在水体边缘地区,有两方面的原因:(1)该类区域内植被对像元散射特征的影响;(2)不同极化方式对水陆边界区域响应不同。

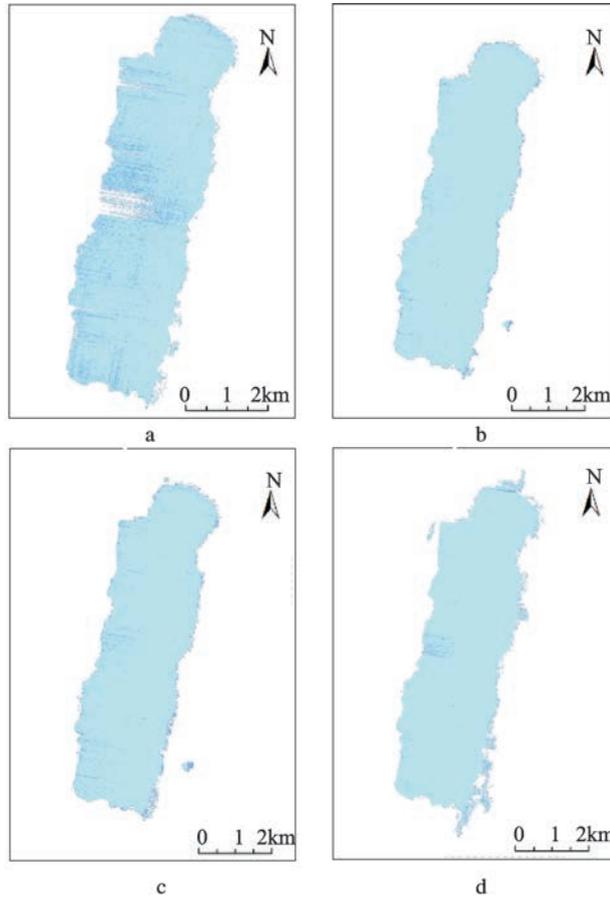


图 4 阳宗海 4 种极化影像提取结果: a HH; b HV; c VH; d VV  
 Fig.4 Extraction results of 4 polarization images in Yangzonghai Lake

表 2 各极化方式影像提取水体面积  
 Table 2 Water area of extraction results in each polarized image

影像	HH	HV	VH	VV	人工标注
面积 (km <sup>2</sup> )	28.82	31.39	31.36	32.42	31.83

表 2 展示了各影像所提水体区域的面积。除 HH 极化外，其他 3 种极化方式所提面积均与在 GF-1 影像上人工标注所得的面积相差不大于 2%，一定程度上反映出水体提取精度对于极化方式的选取并不敏感。

### 3 高分三号业务化水体提取流程

实验证明基于单景雷达影像自动提取水体的理论和方法具有可行性，该方法速度快，不需人工选取阈值，具有较高精度，能满足洪涝监测、水资源管理等业务化需求。

2019年10月

在合适的软、硬件运行环境下，利用该自动化算法实现业务化水体提取的流程如图5所示。

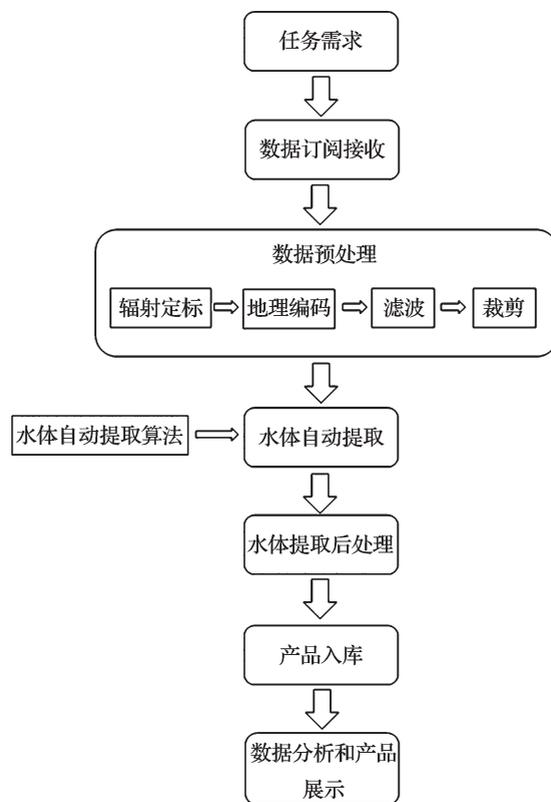


图5 业务化水体提取流程图

Fig.5 Workflow of operational water body extraction

(1) 数据订阅接收：根据任务需求向中国资源卫星应用中心申请订购监测范围内的高分三号影像数据，通过专用光纤接收由中国资源卫星中心推送的影像数据，数据为L1A级产品，即斜距复数影像。

(2) 数据预处理：对获取的数据进行数据预处理，包括多视处理、辐射定标、地理编码、滤波和影像裁剪。根据元信息文件和卫星姿轨参数，经过辐射定标和地理编码，斜距复数影像转换为地距后向散射系数影像。利用滤波技术，消除影像中的斑点噪声，提高信噪比。利用水利普查数据，确定水体的大概范围，以此为基础，采用一定规则确定缓冲区大小，进行影像裁剪，得到一个主要包含水体信息的矩形影像。

(3) 水体自动提取：基于上一步裁剪后的影像，利用雷达水体提取算法进行自动水体提取工作。

(4) 水体提取结果后处理：利用形态学模型进行简单的后处理工作，优化水体矢量数据。

(5) 产品入库：将经过后处理并审核通过的水体提取产品自动归档数据库中，包括

水体的空间矢量数据、面积数据和相关的元数据，根据监测时间，如水文数据库（包括该水体），同步提取水文监测数据，用于后续分析和展示。

（6）数据分析和产品展示：根据业务分析需求，主要完成单个面积变化分析、区域水体面积变化分析、结合水文监测数据的水量变化分析。设计产品展示形式，开发标准服务接口，为各业务应用系统提供接口服务。

## 4 结论

GF-3 卫星为洪涝监测、旱情监测等水资源管理业务提供了有力的数据支撑。基于 GF-3 影像在水体提取中的巨大应用潜力，该文设计了一种基于影像直方图自动化阈值分割的水体提取方法。该方法简单、快速且无需人工干预，适用于业务化水体提取任务。分别从理论和实验两方面验证了该方法的有效性，分析了不同极化方式对水体提取进度的影响，发现 4 种极化方式下的提取精度均能满足水体提取任务的需求，在批量生产中可以根据实际情况任选一种极化方式的数据提取水体。该文根据实际业务需求整理了高分三号业务化水体提取流程，为自动化完成高分三号 SAR 数据从数据接收到获得水体提取结果提供了规范性操作指导，为湖库日常监测、洪涝灾害监测等涉水事件提供了有力的技术支撑，高效的水体自动提取也为农田水利管理提供了技术支持。后续将进一步开展基于极化 SAR 的水体提取研究，提高 GF-3 卫星在水利领域的应用水平。

## 参考文献

- [ 1 ] Oberstadler R., Hunsch H., Huth D. Assessment of the mapping capabilities of ERS-1 SAR data for flood mapping: a case study in Germany. *Hydrological Processes*, 1997 (10): 1415~1425.
- [ 2 ] Roo A. D., Knijff, Horritt M., et al. Assessing flood damages of the 1997 Oder flood and the 1995 Meuse flood. *Proceedings of the Second International ITC Symposium on Operationalization of Remote Sensing, Enschede, The Netherlands*, 1999: 16~20.
- [ 3 ] Townsend P. A. Estimating forest structure in wetlands using multitemporal SAR. *Remote Sensing of Environment*, 2002, 79 (2/3): 288~304.
- [ 4 ] Lv W., Yu Q., Yu W.. Water extraction in SAR images using GLCM and support vector machine. *IEEE 10th International Conference on Signal Processing Proceedings*, 2010.
- [ 5 ] Gong M., Li Y., Jiao L., et al. SAR change detection based on intensity and texture changes. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2014 (93): 123~135.
- [ 6 ] Bates P. D., Horritt M. S., Smith C. N., et al. Integrating remote sensing observations of flood hydrology and hydraulic modelling. *Hydrological processes*, 1997, 11 (14): 1777~1795.
- [ 7 ] Horritt M. S.. A statistical active contour model for SAR image segmentation. *Image and Vision Computing*, 1999, 17 (3/4): 213~224.
- [ 8 ] Otsu N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 1979, 9 (1): 62~66.
- [ 9 ] 李景刚, 黄诗峰, 李纪人. ENVISAT 卫星先进合成孔径雷达数据水体提取研究—改进的最大类间方差阈值法. *自然灾害学报*, 2010, 19 (3): 139~145.
- [ 10 ] Kiani H., Safabakhsh R., Khadangi E.. Fast recursive segmentation algorithm based on Kapur's entropy. *Computer, Control and Communication, IC4 2009, 2nd International Conference on IEEE*, 2009: 1~6.
- [ 11 ] Chini M., Giustarini L., Hostache R., et al. An automatic SAR-based flood mapping algorithm combining hierarchical tiling and change detection. *ESA Living Planet Symp., Prague, Czech Republic*, 2016.
- [ 12 ] Wilks D. S.. Statistical methods in the atmospheric sciences. *Academic press*, 2011.



2019年10月

- [ 13 ] Sahoo P. K., Soltani S., Wong A. K. C.. A survey of thresholding techniques. *Computer vision, graphics, and image processing*, 1988, 41 (2): 233~260.
- [ 14 ] Giustarini L., Hostache R., Kavetski D., et al. Probabilistic flood mapping using synthetic aperture radar data. *IEEE Transactions on Geoscience and Remote Sensing*, 2016, 54 (12): 6958~6969.

## Research on automation and operability in water extraction based on GF-3

Cui Qian<sup>1</sup>, Mao Xudong<sup>2</sup>, Chen Deqing<sup>1</sup>, Qu Meng<sup>2</sup>, Bai Jueying<sup>2</sup>,

( 1. Information Center/Hydrology Monitor and Forecast Center, Ministry of Water Resources, Beijing 100053, China;

2. Wuhan University, Hubei Wuhan 430072, China )

**Abstract:** [ **Purpose** ] Due to its capacity of providing all-time and all-weather observations, automation and operability in water extraction based on GF-3 strongly supports the use of domestic satellite data in operational applications associated with flood monitoring and water resource management. [ **Method** ] This paper introduces an automated water extraction algorithm, which automatically performs threshold segmentation according to the histogram distribution characteristics. For nationwide operational monitoring, this paper presents a processing system containing data subscription reception, data preprocessing, water extraction, afterprocessing, product storage, data analysis and exhibition. [ **Result** ] Yangzonghai Lake, Hongze Lake and Miyun Reservoir are used to validate the effectiveness, and the effect of polarization on extraction result is analyzed. ( 1 ) The histogram exhibits double peak feature. Water body and background pixels are accurately separated by the threshold value. ( 2 ) Polarization has little influence on the water extraction algorithm. [ **Conclusion** ] Water extraction based on GF-3 automatically can effectively monitor water resources and provide data support for the management of farmland water conservancy such as flood and drought.

**Key words:** GF-3; water extraction; automatic threshold; operational method