

## 基于天宫二号多光谱数据的青藏高原冻湖自动提取

刘 康<sup>1</sup>,任海根<sup>2</sup>,李盛阳<sup>1</sup>,覃帮勇<sup>1</sup>

- (1. 中国科学院太空应用重点实验室 中国科学院空间应用工程与技术中心,北京 100094;
2. 中国载人航天工程办公室,北京 100071)

**摘 要:** 高原湖泊在反映全球气候变化背景下区域自然环境变迁方面具有重要意义。以新型国产遥感数据源天宫二号多光谱数据为基础采用面向对象方法,结合水体指数和高程信息,提出一种面向冻湖的自动提取算法。该算法充分考虑了不同形态的水体特性,可以同时提取结冰和未结冰的湖泊,并能够排除冰川、河流的影响。针对选定的 7 个典型区域,采用自动提取算法进行湖泊提取试验,并进行精度验证。湖泊提取整体精度达 99.10%,F-score 为 0.982。结果表明:天宫二号多光谱数据在高原湖泊提取方面具有较强应用潜力,该数据作为一种有效的数据源,可推广用于青藏高原地区湖泊提取与变化研究,为研究区域气候变化提供数据支持。

**关键词:** 青藏高原; 天宫二号多光谱数据; 冻湖提取; 面向对象

中图分类号: TP79 文献标志码: A DOI: 10.3788/IRLA201948.0303004

## Automatic extraction of Tibet Plateau frozen lake based on Tiangong-2 multi-spectral data

Liu Kang<sup>1</sup>, Ren Haigen<sup>2</sup>, Li Shengyang<sup>1</sup>, Qin Bangyong<sup>1</sup>

- (1. Technology and Engineering Center for Space Utilization, Key Laboratory of Space Utilization, Chinese Academy of Sciences, Beijing 100094, China; 2. China Manned Space Agency, Beijing 100071, China)

**Abstract:** Plateau lakes were of great significance in reflecting changes of regional natural environment under the background of global climate change. Based on the new homemade remote sensing data source, Tiangong-2 multi-spectral data, an automatic extraction algorithm for frozen lake was proposed by used object oriented method combined with water index and elevation information. The algorithm took full account of the characteristics of different forms of water, which can simultaneously extract frozen and unfrozen lakes, and exclude the influence of glaciers and rivers. For the selected 7 typical regions, automatic extraction algorithm was used to carry out lake extraction test and accuracy verification. The overall accuracy of lake extraction is 99.10% and F-score is 0.982. The results show that Tiangong-2 multi-spectral data has a significant application potential in the extraction of plateau lakes. As an effective data source, it can be applied to the study of lake extraction and change in the Tibetan Plateau, and provides data support for the study of regional climate change.

**Key words:** Tibetan Plateau; Tiangong-2 multi-spectral data; extraction of frozen lakes; object-oriented

收稿日期: 2018-11-05; 修订日期: 2018-12-13

基金项目: 国家科技部国家科技基础条件平台项目(DKA2018-12-02-23)

作者简介: 刘康(1989-),女,助理工程师,硕士,主要从事环境遥感、图像处理等方面的研究。Email:liukang@csu.ac.cn

## 0 引言

青藏高原平均海拔 4 000 m 以上,面积约 2 500 000 km<sup>2</sup>,被称为“世界屋脊”<sup>[1-2]</sup>。青藏高原分布着地球上海拔最高、数量最多、面积最大的高原湖泊群<sup>[3]</sup>。区域内湖泊受人类活动影响较小,绝大部分处于自然状态,可为气候变化研究提供重要指标,是气候变化的前哨<sup>[4]</sup>。近年来,很多学者开展青藏高原湖泊变化的研究,特别是基于遥感监测的青藏高原湖泊提取与变化。相较于传统的湖泊监测方法,利用遥感技术能更全面、更快速地获取湖泊的动态变化,尤其在自然环境恶劣、缺少连续地面观测数据的青藏高原高寒地区。目前,基于遥感数据的水体分类方法主要包括四类:专题分类法、线性混合模型法、单波段阈值法和光谱水体指数法<sup>[5]</sup>。其中,归一化差异水体指数法(Normalized Difference Water Index, NDWI)是应用最广泛的方法<sup>[6]</sup>。目前很多学者也提出了其他水体指数,例如改进的归一化差异水体指数、AWEI 等<sup>[7-9]</sup>。随着遥感应用技术的不断发展,自动分类方法为高原湖泊提取提供了新的机遇和挑战,例如面向对象的图像分析技术<sup>[10]</sup>。这种自动进行地理信息提取的方法将图像分割呈“像元斑”对象,避免了椒盐现象的产生,不仅可以对大区域进行快速调查,还可以减少手工数字化产生的操作者偏差<sup>[11-12]</sup>。但是,对于青藏高原地区,单一的指数和阈值往往把湖泊与冰川、云覆盖以及高反射的沉积物混淆起来,因此,基于面向对象的分析算法需要更为精细化的设计。

高原湖泊研究常用的数据源有美国国家航空和航天局 Landsat 数据(1972 年~至今),包括 Landsat MSS(1972~1992 年)、TM(1982 年)、ETM+(1999 年)和 Landsat-8OLI(2013 年)。此外,其他更多中-高分辨率遥感数据(如 MODIS、Sentinel 等)也被用于青藏高原湖泊提取研究。天宫二号空间实验室搭载的宽波段成像仪具有宽刈幅、高时间分辨率等特点,并且其谱段设置对水色较为敏感,将成为青藏高原湖泊提取研究方面的又一新的数据源,但目前尚无该领域的研究。

文中选取天宫二号宽波段成像仪影像为数据源,利用面向对象的提取方法并结合多种水体指数,开展高原湖泊提取研究。文中研究论证了天宫二号

多光谱数据在高原湖泊提取方面的应用潜力,该数据可作为一种有效的数据源,广泛用于青藏高原地区湖泊提取与变化研究,为研究区域气候变化提供数据支持。

## 1 研究区概况

### 1.1 研究区

青藏高原(26°00′~39°47′N,73°19′~104°47′E)南起喜马拉雅山脉南缘,北至昆仑山、阿尔金山和祁连山北缘,西部为帕米尔高原和喀喇昆仑山脉,东及东北部与秦岭山脉西段和黄土高原相接,总面积约 2 500 000 km<sup>2</sup>,平均海拔 4 000~5 000 m,是中国最大、世界海拔最高的高原<sup>[1-2]</sup>。青藏高原上分布着海拔最高、面积最大、数量最多的湖泊群<sup>[3]</sup>。其中,面积大于 1 km<sup>2</sup> 的湖泊有 1 200 多个,约占中国湖泊总数量和面积的 50%<sup>[4]</sup>。湖泊的分布情况如图 1 所示。

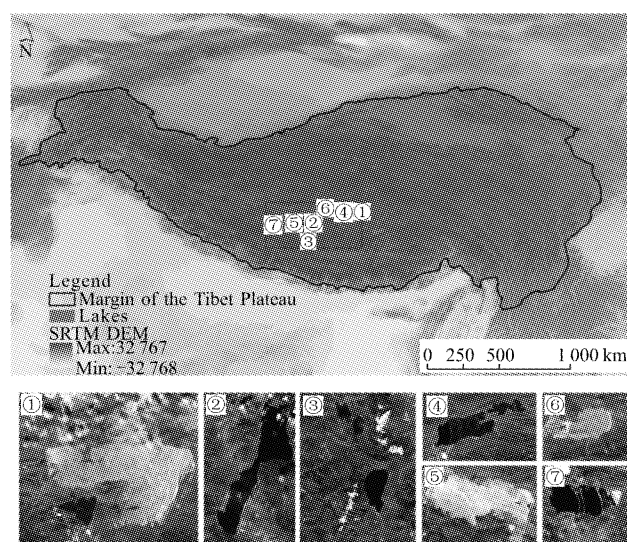


图 1 青藏高原及典型研究区位置分布图

Fig.1 Distribution of Tibetan Plateau and its typical research areas

### 1.2 数据源

该研究主要数据源为天宫二号宽波段成像仪获取的多光谱数据(MWI,获取于载人航天空间应用数据推广服务平台,网址 <http://www.msadc.cn/>)。天宫二号宽波段成像仪具有可见近红外(VNI)、短波红外(SWI)和热红外(INF)多光谱大视场全推扫成像的组合集成功能。影像幅宽为 300 km,其中宽波段成像仪可见近红外谱段空间分辨率是 100 m,具有 14 个通道,光谱范围是 0.4~1.0 μm;短波红外谱段空间分

分辨率是 200m,具有两个通道,光谱范围是 1.0~1.7 $\mu\text{m}$ ;热红外谱段空间分辨率是 400 m,具有 2 个通道,光谱范围是 8~10  $\mu\text{m}$ <sup>[13]</sup>。

## 2 研究方法

基于天宫二号多光谱数据,文中利用面向对象方法并结合水体指数和高程信息,提出一种面向冻湖提取的算法。这套算法充分考虑了不同形态的水体特性,可以同时提取结冰和未结冰的湖泊,并能够排除冰川的影响。首先计算 NDWI 指数和雪盖指数(Normalized Difference SnowIndex, NDSI),然后针对指数计算结果进行分割,并结合坡度、形状特征参数等信息对分割后的对象属性特征进行分析,设计提取阈值,分别对 7 个典型区域进行湖泊自动提取实验,最后验证结果精度。具体的自动提取流程如图 2 所示。

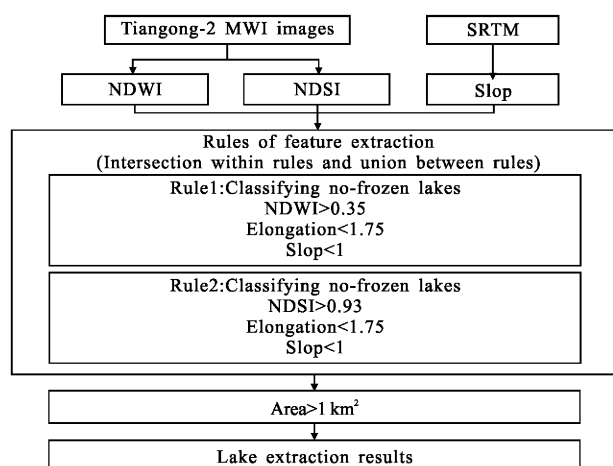


图 2 面向对象提取过程图

Fig.2 Object-oriented procedure chart of the extraction

### 2.1 影像预处理

天宫二号宽波段成像仪二级数据是经过辐射校正和几何校正的产品,因此只需对其进行大气校正。文中采用 ENVI5.3 中的 FLAASH 模型对影像进行大气校正,其中 Radiance Scale Factors 选择 Use single scale factor for all bands (Single scale factor=100), Sensor Altitude=390 km。采用 ENVI5.3 Registration 工具将可见近红外和短波红外图像进行配准,精度在 0.5 个像元以内。最后,将可见近红外和短波红外数据进行波段组合,合成的影像波段数为 16 个。

### 2.2 水体指数和坡度计算

#### 2.2.1 水体指数

从众多的水体指数中,充分考虑指数特点以及数据情况,文中选择 NDWI 和 NDSI 两个指数<sup>[16,14]</sup>如表 1 所示。

表 1 水体指数  
Tab.1 Water index

Index	Equation	Author
NDWI	$NDWI = \frac{Green - NIR}{Green + NIR}$	McFeeters, 1996
NDSI	$NDSI = \frac{Green - SWI}{Green + SWI}$	Hall, 1995

注: 天宫二号多光谱数据 Green =band10, NIR =band4, SWI =band16; Landsat TM: Green=band2, NIR=band4, SWI=band5

水体的反射率在绿光波段高于近红外波段,而土壤和植被则恰好相反,采用 NDWI 指数可以增强水体信息的同时抑制土壤和植被等非水体信息<sup>[12]</sup>。而冻湖以及冰川、雪山等地物绿光波段的反射率较近红外波段低,其 NDWI 值低于非冻湖。若 NDWI 分割阈值偏高,导致冻湖部分缺失,造成欠分割的现象;若 NDWI 分割阈值偏低,导致 NDWI 值较低的冰川、山体阴影被划分为湖泊,造成过度分割现象<sup>[12]</sup>。单一指数无法同时提取冻湖和非冻湖,因此文中采用 NDSI 提取冻湖。NDSI 将可见光的强反射和短波红外低反射进行归一化处理,突出冰雪特征。

#### 2.2.2 坡度

选择 SRTM(90 m)数据计算坡度信息(数据来源 <http://srtm.csi.cgiar.org/>)。将数据进行拼接处理,重采样成 100 m 分辨率(与多光谱数据保持一致),利用 Topographic Modeling 计算坡度。

### 2.3 空间特征提取模型参数选择与规则设置

冻湖提取采用面向对象的图像分析方法。不同于传统的基于每个像素的光谱信息来进行图像分类的方法,面向对象的方法是先将图像分割成光谱信息、纹理特征等相似的对象,再对对象进行图像分类。过程中涉及对象的多种属性,如光谱特征、形状特征、纹理特征等。面向对象的方法对于分类的精度和效率都有所提高。文中研究采用 ENVI 5.1 面向对象空间特征提取模块(ENVI Feature Extraction-FX)。

#### 2.3.1 参数选择

模型的参数包括分割系数和合并系数。参数选

择对提取精度有一定影响。为了选择合适的参数,以区域⑤为例,通过调整不同的参数组合,提取湖泊,并将结果与参考数据(目视解译的结果)进行对比,差值越小,表明阈值越合理。

定义  $d$  为基于面向对象提取结果与参考数据集的差值的平方根,用来反映参数设置对于提取结果的好坏程度,值越小表示参数设置越合理。

$$d = \sqrt{a' - a} \quad (1)$$

式中: $a'$ 为面向对象提取结果的湖泊面积; $a$ 为参考数据的湖泊面积。

### 2.3.2 提取规则设定

设置特征提取的规则,首先针对 NDWI 指数设置提取阈值,提取未结冰的水体,其次,针对 NDSI 设置提取阈值,将结冰和积雪的区域提取出来。为了避免河流的干扰,通过形状参数 Elongation 排除长条形的对象;为了避免冰川的干扰,通过坡度信息,去除坡度大于  $1^\circ$  的区域。

### 2.4 精度验证

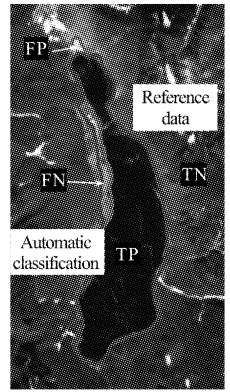
为了检验结果的精度,文中选择 8 个评价指标来评价结果的精度,分别是:误差 1(Type1 error)、误差 2(Type2 error)、整体误差(Total error)、整体精度(Overall accuracy)、产品精度(Producer's accuracy)、用户精度(User's accuracy)、Kappa 系数(Cohen's kappa)和 F 得分(F-score)(表 2)。参考数据集通过目视解译

获得。最终通过上述指标评价错分(提取的非湖泊区域)和漏分区域(未提取湖泊区域)情况。

表 2 湖泊提取结果精度评价指标

Tab.2 Evaluation index of lake extraction precision

Index	Equation
Type1 error	$FP/(TP+FP)$
Type2 error	$FN/(FN+TN)$
Total error	$(FP+FN)/(TP+FP+FN+TN)$
Overall accuracy	$(TP+TN)/(TP+FP+FN+TN)$
Producer's accuracy	$TP/(TP+FP)$
User's accuracy	$TP/(TP+FN)$
Cohen's kappa	$(Po-Pe)/(1-Pe)$
F-score	$2TP/(2TP+FP+FN)$



注:TP 代表湖泊区域;FP 代表漏分区域;FN 代表错分区域;TN 代表非湖区域

## 3 结果与分析

### 3.1 指数计算结果

基于 NDWI 和 NDSI 指数的计算结果如图 3 所示。图中深至浅,表示指数值由高到低。NDWI 指数可以增强水体信息,同时抑制土壤等非水体信息。未结冰的水体 NDWI 值较高,结冰或被积雪覆盖的湖

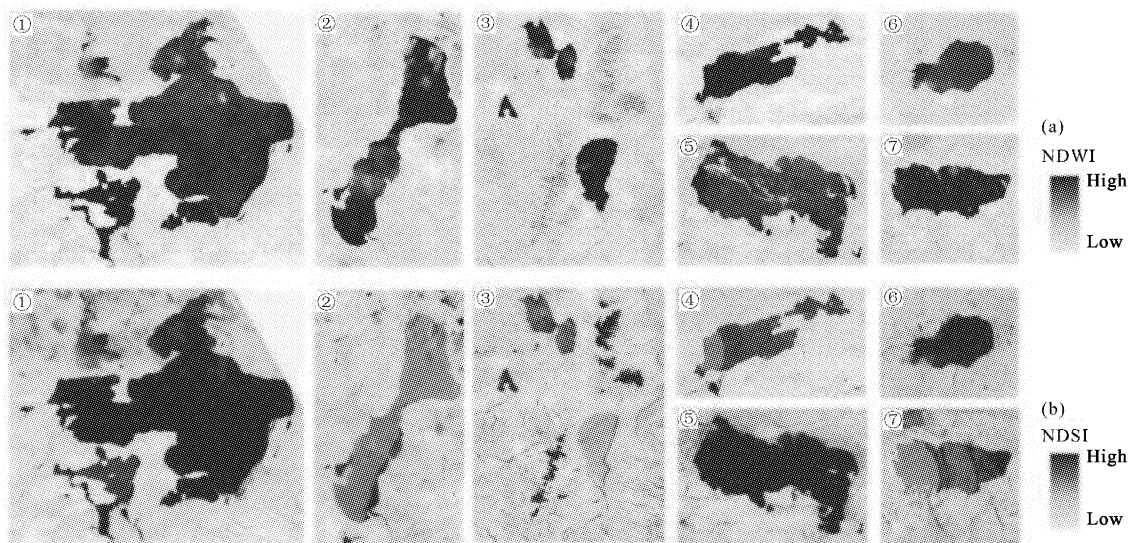


图 3 NDWI 和 NDSI 指数计算结果

Fig.3 Calculation results of NDWI and NDSI

面值较低。此外,NDWI 被云遮挡的湖面值较低,与非水体接近。

NDSI 中采用短波红外波段,对冰雪较为敏感,且对云雾有一定的穿透能力。NDSI 指数计算结果中,被冰、雪覆盖的区域值较大,冰雪信息被增强。

### 3.2 指数分割阈值选择

图 4 是 7 个区域 NDWI 和 NDSI 指数的统计结果图。图 4 中 NDWI 的统计图中呈现显著双峰,两峰

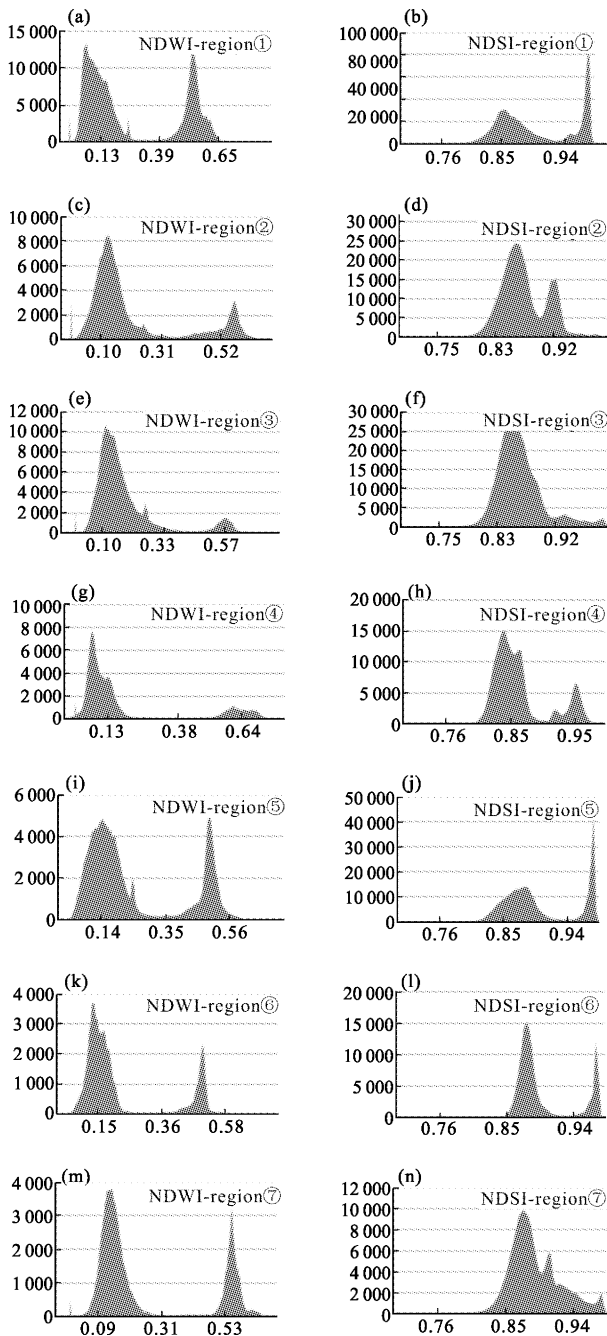


图 4 7 个典型区域 NDWI 和 NDSI 指数统计结果

Fig.4 NDWI and NDSI statistics results of seven typical regions

较为离散。值较小的区域为非水体,值较大的区域为水体。两峰之间为分割阈值,但是不同区域,分割阈值有一定差别。如果 NDWI 分割阈值偏高时,导致冰冻湖泊部分缺失,造成欠分割的现象;如果 NDWI 分割阈值偏低时,导致 NDWI 值较低的冰川、山体阴影被划分为湖泊,造成过度分割现象。因此,单一指数无法将湖泊提取出来。文中研究选择 0.35,将未结冰湖面提取出来,结冰湖面通过 NDSI 指数结果进行插补。

NDSI 的统计结果双峰现象不明显,这是由于 NDSI 将冰雪区域和非冰雪区域分开,选择的 7 个区域中,两区域分布的像元数有悬殊。文中研究选择 0.93,将被积雪覆盖的区域提取出来。

### 3.3 面向对象的提取结果

#### 3.3.1 分割阈值和合并阈值选择

分割参数 (Scale Level,SL) 和合并参数 (Merge Level,ML) 阈值的选择对结果有较大的影响,选择最佳的阈值有助于提高分类的准确性。图 5 是不同分割参数和合并参数组合下,研究区域⑤的分割情况:分割参数由小到大呈现由过分割到欠分割的过度;合并参数由小到大呈现由过分割到欠分割的过度。

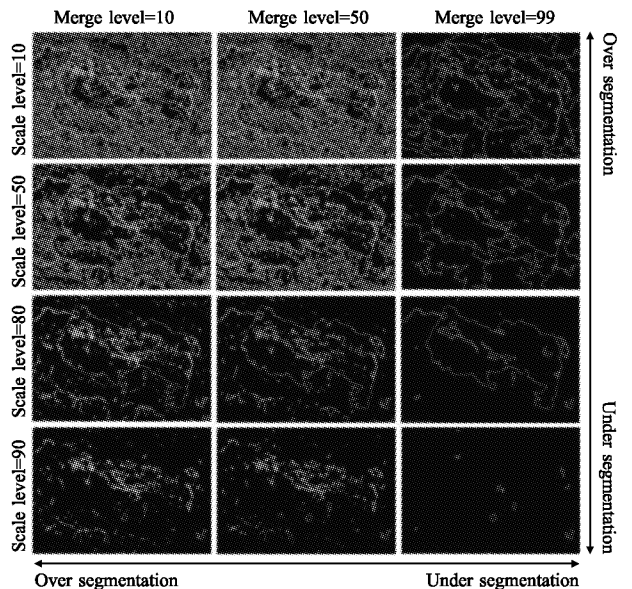


图 5 区域⑤分割结果

Fig.5 Segmentation results of region⑤

笔者将两个参数由 5~95、步长为 5,分别进行湖泊提取,并统计湖区面积。用参数  $d$  衡量提取结果的好坏,值越小则表明对应的参数选择越合理。由图 6

可知,  $SL > 85$  或者  $ML = 100$  时,  $d$  值出现突变, 提取效果极差;  $SL = [30, 50]$  且  $ML \leq 100$  时,  $d$  值较小, 且较为接近, 表明  $SL$  对提取结果的影响较大, 而  $ML$  对结果影响较小。当  $SL = 45$  且  $ML = 95$  时, 提取面积与真实值最接近。后续的提取过程选择该阈值进行影像分割。

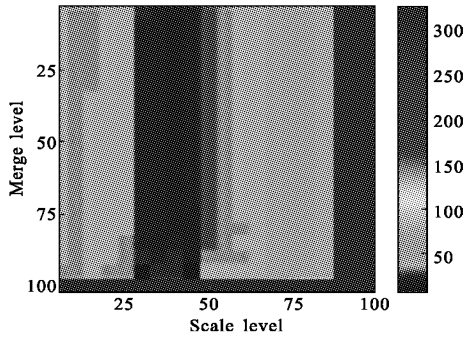


图 6 面向对象分割阈值和合并阈值选择

Fig.6 Selection of object-oriented segmentation threshold and merge threshold

### 3.3.2 提取结果与精度检验

基于天宫二号多光谱数据, 采用文中提出的双指数面向对象的提取算法, 结果如图 7 所示。图中红色矢量线围绕的区域为提取的湖区。由图 7 可知, 提取结果与湖泊区域基本一致, 结冰和未结冰的湖泊均被提取出来, 河流等线性水体(区域⑤)以及冰川(区域③)均被剔除。少数漏分的区域为云遮挡的湖区(区域②)。

通过目视解译, 并结合相近时相的 Landsat OIL 影

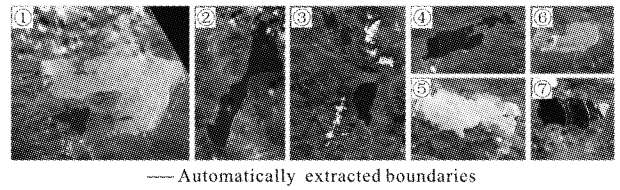


图 7 面向对象提取结果

Fig.7 Results of object-oriented extraction

像数据, 提取研究区湖泊的矢量边界作为参考数据。将基于面向对象的提取结果与参考数据对比, 计算精度评价结果, 如表 3 所示。

由 3 表可知, 文中提出的方法精度较高, 整体体精度达到 0.991, 产品精度和用户精度分别为 0.984 和 0.981, F 得分为 0.982。

7 个典型区域提取精度有一定差异, 区域①、③、⑤、⑦的提取精度较高, 大于均值。区域②、④、⑥的提取精度较低, 将几个区域漏分、错分的结果放大展示, 如图 8 所示。

由图 8 可以看出, 区域②中 A、B 两个区域被云覆盖, 造成漏分; C 区域的高亮区被误认为水体, 造成错分。表 3 统计结果显示, 区域②的提取精度最差, 且误差 1 达到 0.045 3, 表明漏分较多, 这是由于区域②被云覆盖的区域未被提取出来。区域④中 A、B 两个高亮区域被误认为水体, 造成错分, 因此 C 处出现少许漏分现象。同时, 区域⑥中 A 处存在水面较浅、湖底裸露的区域, 且未被提取出来, 出现漏分现象。

表 3 精度检验

Tab.3 Accuracy verification

	Type1 error	Type2 error	Total error	Overall accuracy	Producer's accuracy	User's accuracy	Cohen's kappa	F-score
Region①	0.012	0.005	0.007	0.993	0.988	0.991	1.003	0.990
Region②	0.045	0.013	0.020	0.980	0.955	0.953	1.608	0.954
Region③	0.019	0.001	0.003	0.997	0.981	0.986	1.004	0.984
Region④	0.017	0.011	0.012	0.988	0.984	0.957	0.982	0.970
Region⑤	0.006	0.007	0.007	0.993	0.995	0.985	0.985	0.990
Region⑥	0.015	0.010	0.011	0.989	0.985	0.968	0.988	0.976
Region⑦	0.014	0.004	0.007	0.993	0.986	0.991	0.992	0.989
Average	0.017	0.007	0.009	0.991	0.984	0.981	1.000	0.982

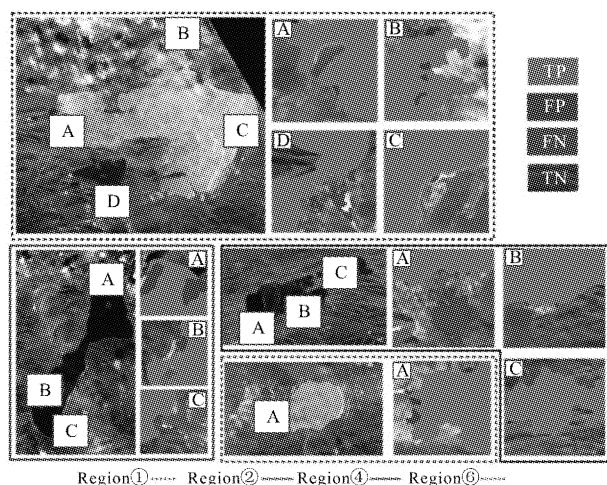


图 8 典型区域漏分错分结果图(区域①、②、④、⑥)

Fig.8 Leakage and misclassification results of typical regions (Region①, ②, ④, ⑥)

## 4 结 论

文中选取天宫二号多光谱数据为数据源, 开展青藏高原冻湖提取研究, 主要结论如下:

(1) 基于天宫二号多光谱数据的 NDWI 和 NDSI 指数虽然波段选取与 Landsat TM 数据有差异, 但提取结果良好;

(2) 面向对象提取结果与分割参数和合并参数选择有关, 并且分割参数对结果影响更大。分割参数在 30~50 之间, 合并参数小于 100 时, 分割结果较好;

(3) 文中提出的基于面向对象的双指数冻湖自动提取方法可以提取表面结冰和未结冰的湖泊, 并且避免河流、冰川等的影响, 总体精度达 99.10%, F 得分为 0.982。表明天宫二号多光谱数据在青藏高原冻湖提取方面具有明显优势;

(4) 文中研究有云遮挡的区域的提取精度较低, 为了减少云遮挡对结果的影响, 可考虑选择相近时相的数据通过时间插补的方法提高湖泊提取精度。

## 参 考 文 献:

[1] Qiu J. The third pole [J]. *Nature*, 2008, 454(7203): 393-396.  
 [2] Wang C. A remote sensing perspective of alpine grasslands on the Tibetan Plateau: Better or worse under "Tibet Warming"? [J]. *Remote Sensing Applications Society &*

*Environment*, 2016, 3: 36-44.

[3] Che X, Feng M, Jiang H, et al. Detection and analysis of qinghai-tibet plateau lake area from 2000 to 2013 [J]. *Journal of Geo-Information Science*, 2015, 17(1): 99-107.  
 [4] Zhang G. Changes in lakes on the Tibetan Plateau observed from satellite data and their responses to climate variations [J]. *Progress in Geography*, 2018, 2: 214-223.  
 [5] Ji L, Zhang L, Wylie B. Analysis of dynamic thresholds for the normalized difference water index [J]. *Photogrammetric Engineering & Remote Sensing*, 2009, 75(11): 1307-1317.  
 [6] 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.  
 [7] Xu H. 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.  
 [8] Feyisa G L, Meilby H, Fensholt R, et al. Automated water extraction index: a new technique for surface water mapping using landsat imagery [J]. *Remote Sensing of Environment*, 2014, 140(1): 23-35.  
 [9] Zhang G, Li J, Zheng G. Lake-area mapping in the Tibetan Plateau: An evaluation of data and methods [J]. *International Journal of Remote Sensing*, 2017, 38 (3): 742-772.  
 [10] Blaschke T. Object based image analysis for remote sensing [J]. *ISPRS Journal of Photogrammetry & Remote Sensing*, 2010, 65(1): 2-16.  
 [11] Korzeniewska K, Korup O, Korzeniewska K, et al. Object-based detection of lakes prone to seasonal ice cover on the tibetan plateau [J]. *Remote Sensing*, 2017, 9(4): 339.  
 [12] Huang T J, Liang D D, Jia L, et al. Automatic extraction of lake area with iterative interpolation method on tibetan plateau [J]. *Remote Sensing Technology and Application*, 2017, 32(2): 289-298.  
 [13] Qin B Y, Li S Y, Hao Z W, et al. Earth observation data of the Tiangong -2 Space Laboratory (2016 -2017) [J/OL]. [2017-12-20]. <http://www.10.11922/csdata.2017.0003.zh>.  
 [14] Hall D K, Riggs G A, Salomonson V V. Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data [J]. *Remote Sensing of Environment*, 1995, 54(2): 127-140.