文章编号: 0253-374X(2011)08-1247-06

DOI:10.3969/j.issn.0253-374x.2011.08.028

基于遥感影像光谱分析的蓝藻水华识别方法

林 怡^{1,2,3},潘 琛^{1,2},陈映鹰^{1,2},任文伟^{3,4}

(1.同济大学 遥感与空间信息技术研究中心,上海 200092;2.同济大学 测量与国土信息工程系,上海 200092;3.中国-加拿大环境与可持续发展中心,上海 200433;4.复旦大学 生命科学学院,上海 200433)

摘要:利用 Landsat-7 ETM+遥感影像数据,以淀山湖为例, 在分析蓝藻和其他典型地物影像光谱曲线及其特征的基础 上,构建归一化蓝藻指数(NDI_CB),有效地从浑浊水体中提 取蓝藻信息.通过 k-均值非监督分类结果可以发现,构建的 归一化蓝藻指数较传统的归一化差值植被指数(NDVI)和比 值植被指数(RVI)更加适用于提取低密度蓝藻空间分布信 息.在此基础上,基于遥感影像光谱特征和归一化蓝藻指数, 采用了支持向量机的分类识别模型,最终得到淀山湖区域蓝 藻的空间分布范围与面积,通过发现在某一特定时间蓝藻分 布的规律,为蓝藻预警和治理的生态学分析提供了及时、有 效和客观的依据.

关键词: 归一化蓝藻指数;光谱分析;蓝藻识别;支持向量机 分类

中图分类号: TP 751.1; TP 79 文献标识码: A

Recognition of Cyanobacteria Bloom Based on Spectral Analysis of Remote Sensing Imagery

LIN Yi^{1,2,3}, PAN Chen^{1,2}, CHEN Yingying^{1,2}, REN Wenwei^{3,4} (1. Research Center of Remote Sensing and Spatial Information Technology, Tongji University, Shanghai 200092, China; 2. Department of Surveying and Geoinformatics, Tongji University, Shanghai 200092, China; 3. Sino-Canada Centre for Environment and Sustainable Development, Shanghai 200433, China; 4. School of Life Sciences, Fudan University, Shanghai 200433, China)

Abstract: Based on the analysis of spectral curve and features of cyanobacteria bloom and other typical ground object, the normalized difference cyanobacteria bloom index(NDI_CB) was constructed to distinguish between cyanobacteria bloom and turbid water with the Landsat - 7 ETM + image in Lake Dianshan. In this study two other different vegetation indexes, normalized difference vegetation index(NDVI) and ratio vegetation index(RVI), together with NDI_CB, were applied to extracting the cyanobacteria bloom information from the same image via unsupervised classification method (kmeans). The results show that NDI_CB is the best one for low-density cyanobacteria bloom extraction. In order to recognize the cyanobacteria bloom better, support vector machine(SVM) classification method was used to classify the image based on spectral features and NDI_CB, and to obtain the spatial distribution and the area of cyanobacteria bloom in Lake Dianshan. Through studying the laws of the cyanobacteria bloom distribution at a particular time, a sound, efficient and objective basis has been achieved for the ecological analysis of the prevention and the treatment of cyanobacteria bloom.

Key words: normalized difference cyanobacteria bloom index; spectral analysis; recognition of cyanobacteria bloom; support vector machine classificatin

近年来,由于内陆湖泊富营养化严重,导致蓝藻 水华频繁发生,不仅恶化了水体的感官性状,而且破 坏了湖泊生态系统的总体平衡,严重降低了湖泊的 使用功能,给居民生活、经济发展等带来严重影 响^[1].针对蓝藻进行治理或控制,必须了解实际环境 下蓝藻的生理生化特征及蓝藻种群的时空动态变化 规律,从而有针对性地实施各种治理方案.

传统的人工现场采集水样进行水质分析的方法 监测结果较准确,但采样频次有限、成本高、耗时长, 对于大型浅水湖泊难以获得整个水体的水质情 况^[2].作为信息化和数字化的重要技术手段,遥感技 术能够提供同一地区多时相、多尺度、多光谱、多平 台的观测信息,通过对影像数据的分析与处理,能够 快速、有效地识别蓝藻及其空间分布,对分析蓝藻暴

收稿日期:2010-12-09

基金项目:科技部国际科技合作计划(2009DFA92310)

第一作者:林 怡(1970—),女,副研究员,博士生副导师,工学博士,主要研究方向为遥感与摄影测量、遥感技术应用.

E-mail: linyi@tongji.edu.cn

发的动态过程有重要意义[3-4].目前,国内外学者利 用遥感技术对蓝藻水华的研究取得了一定的进展. 谢国清等^[5]通过研究滇池蓝藻水华在可见光、红外 谱段的光谱特征并利用假彩色合成法以及归一化植 被指数法进行了滇池蓝藻水华信息的遥感识别和提 取. Anu Reinart 等^[6]结合生物 - 光学模型,比较了 SeaWiFS, MODIS/Aqua 和 MERIS 影像数据在识别 波罗的海部分水域蓝藻水华中的效果,结果表明,在 蓝藻水华暴发的早期,各种影像数据给出的识别结 果较为一致,然而到了水华暴发的严重期,高光谱分 辨率影像 MERIS 和 MODIS 表现出较好的蓝藻水华 识别能力. Richard H. Becker 等[7] 基于 MODIS 影像 数据提取了10景北美洲五大湖部分研究区2004年 和2005年夏季蓝藻空间分布图与实地测定数据检 核得出这种方法提取的结果准确有效,并且能有效 控制成本.

本文使用 2000 年 8 月 1 日获取的 Landsat-7 ETM+遥感数据,试图通过分析淀山湖蓝藻覆盖区 的光谱曲线及其特征,综合更多波段的谱间关系构 建有效的蓝藻指数,并获取淀山湖蓝藻的空间分布 格局,为以后开展长期的淀山湖蓝藻水华动态时空 变化监测与预警提供技术参考和科学依据.

1 蓝藻覆盖区的光谱特征及其分析

1.1 研究区概况与遥感影像数据

淀山湖位于江、浙、沪交界处,地理位置为 120°54′121°01′E,30°02′~31°04′N,总面积约为 62 km²,平均水深2.1 m^[8].它是上海地区唯一的天 然淡水湖泊,水质的好坏直接影响到上海市民的饮 用水质量.

蓝藻作为一种水生浮游藻类在有机氮、有机磷丰 富、无风晴好的天气环境下暴发,淀山湖区域多发生在 7~8月.综合考虑影像数据的空间分辨率、光谱分辨 率、时间分辨率等几项指标要求,并且在经济上要可行, 本文选取 2000 年 8 月 1 日获取的 Landsat-7 ETM+数 据,空间分辨率 30 m,选用 6 个波段(1~5 和 7 波段), 成像质量和成像时的天气状况都比较理想.在进行植被 指数提取与蓝藻覆盖区识别之前,首先利用当日 ETM +影像每个波段的增益和偏置,将数字图像的亮度值 (DN)转化为辐射率,以得到传感器入口处的辐射值.图 1 为 5,4,3 波段假彩色合成图和蓝藻实地拍摄图.

1.2 蓝藻覆盖区光谱特征及其分析

蓝藻暴发时绿色的藻类生物体聚集于水体表面,



Fig. 1 Colorized ETM+ Image(Band 543-RGB) and photograph of cyanobacteria bloom

在外力(如风、水流等)的作用下,大面积积聚,并在 遥感影像上产生类似于陆生植被的光谱特征^[9].以 往研究表明^[5,9-10]:包含蓝藻水华水体的反射光谱在 蓝波段和红波段的吸收峰较为明显,而在近红外波 段具有类似于植被光谱曲线特征的"陡坡效应";在 近红外波段的反射率明显高于水体,是反映蓝藻水 华的主要波段.但是当蓝藻浓度较低时,则表现出与 高悬浮物含量水体相似的光谱特征,给水华面积和 空间分布的提取带来一定的困难.因此,需要对蓝藻 和其他地物的光谱特征进行进一步分析,旨在得到 有效的谱间关系和波段组合来提取水华信息.

在 Landsat-7 ETM + 影像上根据实地采样数据 选取蓝藻、水体(浑浊)和陆生植被为训练样本,光谱 曲线及其特征(影像亮度值)如图 2.



Fig. 2 Spectral curve and character of typical ground objects in ETM+

从图2可以看出:

(1) 蓝藻水华因为混有背景水体的影响,光谱 趋向两者的混合叠加,近红外处的"陡坡效应"也较 陆生植被有所减缓.在波段4处,蓝藻和浑浊水体光 谱特征差异显著,因而可以采用单波段4设立阈值 提取蓝藻.然而阈值的大小直接影响蓝藻的提取,并 且易与高浑浊水体相混淆,高悬浮物水体常被误作 为蓝藻水华. (2)综合考虑波段4和波段3发现,蓝藻水华的 影像亮度值从波段3到波段4逐渐提高,而浑浊水 体则恰好相反,因此可以通过计算波段4与波段3 的影像亮度值的比值是否大于1判断是否为蓝藻水 华.然而该判别式比较适合于水华密度较大的区域, 当密度很小、水体掩盖蓝藻信息时,近红外波段反射 率较低,导致近红外波段和红光波段的影像亮度值 比值小于1,从而低估了水华的集聚和覆盖 面积^[9,11-12].

(3)综合考虑波段3,4和5三个波段发现,蓝藻 水华的亮度值从波段4到波段5的下降程度明显高 于浑浊水体.如果将波段4、波段3和波段5、波段4 分别进行差值处理(如图3所示,图中a,b分别表示 波段4与3的亮度差值、波段5与4的亮度差值), 那么发现蓝藻水华和浑浊水体的趋势截然相反,并 且当蓝藻浓度很低时,由于水体信息的影响,这种趋 势表现得更加显著.



2 归一化蓝藻指数的构建

将某种植被指数作为分类特征参与监督或非监督分类,实现蓝藻水华的识别已经应用于很多湖泊的研究中.刘建萍等^[2]在太湖蓝藻水华遥感监测研究结果的基础上总结分析了植被指数用于蓝藻水华识别的方法,并讨论了现有研究中存在的问题.段洪涛等^[10]利用单波段和谱间关系判别式分别从 MODIS/Terra,ETM和IRS-P6 LISS3遥感数据中提取蓝藻水华信息.彭文祥等^[13]基于多数据源、多时 相卫星数据,选用近红外波段与红波段的相对反射 率之比所生成的图像进行非监督分类处理,并反复 与假彩色图像进行对比分析以提取蓝藻信息.

以往许多植被指数的建立都是针对陆生植被而 言,最常使用波段4和波段3建立的归一化差值植被指 数(normalized difference vegetation index,NDVI)和比值 植被指数(ratio vegetation index,RVI)在蓝藻密度较小 时,由于近红外波段反射率降低,从而导致不能有效地 识别水华.因此,必须对传统植被指数进行改进,建立归 一化蓝藻指数^[5],以达到抑制背景水体及泥沙的影响, 使其得到的结果不论是分布范围还是蓝藻强度都能较 为真实地反映蓝藻水华情况.

根据以上对蓝藻覆盖区和其他典型地物的光谱 特征分析,综合考虑波段 3,4 和 5 三个波段的谱间 关系,本文试图对植被指数进行改进,构造归一化蓝 藻指数 N_{DLCB} (normalized difference cyanobacteria bloom index,NDI_CB),其方法如下:

(1)分别对波段4、波段3和波段5、波段4进行 差值运算.

$$\begin{cases} a = B_4 - B_3 \\ b = B_5 - B_4 \end{cases}$$
(1)

式中, B_3 , B_4 和 B_5 分别表示波段3,4和5经过影像 传感器定标后的光谱值.

(2)由于式(1)得到的差值(a 和 b)跨越正负值 域,为了保证下面的归一化处理能够得到 - 1~1之 间的分布,并且不改变原来直方图的形状,取 a 和 b 中的最小值,分别将 a, b 向上平移最小值个单位, 从而使 a 和 b 全部分布在正值区域.

$$\begin{cases} c = \min(a, b) \\ a' = a + |c| = a + |\min(a, b)| \\ b' = b + |c| = b + |\min(a, b)| \end{cases}$$
(2)

(3) 将式(2)中得到的差值(a'和b')作比值运算,构建归一化蓝藻指数 N_{DI CB}.

$$N_{\rm DLCB} = \frac{a'/b'-1}{a'/b'+1} = \frac{a'-b'}{a'+b'}$$
(3)

得到淀山湖区域影像 NDI_CB 的提取结果,如 图 4 所示.



图 4 研究区 NDI_CB 指数提取结果

Fig.4 Extraction result of NDI_CB in Lake Dianshan 对传统的 NDVI, RVI 和本文构建的 NDI_CB 三种 植被指数分别进行 *k*-均值非监督分类处理,类别数

设定为3(浑浊水体、较高密度覆盖区和低密度覆盖 区),将不同密度覆盖区分类结果进行合并,结果见图 5.为了便于比较和分析,现将低密度覆盖区、较高密 度覆盖区和浑浊水体的影像假彩色合成图与植被指 数(NDI_CB和NDVI)分类图进行分区域显示,如图6 所示.从图5和图6中可以看出,RVI基本上没有提取 出蓝藻水华的空间分布,这是因为研究区该时间点暴发的蓝藻密度不是特别大,也说明了 RVI 不适合提取没有完全覆盖水面的低浓度蓝藻; NDI_CB 提取的蓝藻覆盖区面积较 NDVI 更大些,其中多提取的部分经过假彩色图像比对,很多为低浓度蓝藻覆盖区,但是也会将蓝藻附近的水体误作为蓝藻类别.



图 5 研究区 3 种植被指数 k -均值非监督分类





图 6 假彩色 ETM + 影像与 NDI_CB, NDVI 分区显示

Fig.6 Subareas of colorized ETM+ image and classification images based on NDI_CB or NDVI

3 基于 NDI_CB 的蓝藻支持向量机分 类模型

支持向量机(SVM)^[14]方法应用于分类时具有 适用于高维特征空间、小样本统计学习、抗噪声影响 能力强等特点.由于现场采样数据的局限以及遥感 影像数据空间分辨率不高造成大量混合像元的存 在,能够选取的蓝藻覆盖区训练样本较小,因而采用 SVM 分类方法解决蓝藻和浑浊水体的分类问题.下

1250

面以蓝藻分类识别过程为例,论述基于 N_{DLCB}的蓝藻 SVM 分类模型的理论与规则.

将 N_{DLCB} 与影像传感器定标后各波段光谱值组 合作为分类特征向量 $\mathbf{x} = [b_m; N_{\text{DLAB}}]$,其中, b_m 为 第 m 波段的影像光谱值, $m = 1, 2, \dots, 6$. 给定训练 样本集 $T = \{(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_l, y_l)\} \in (\mathbf{x} \times y)^l$,式 中,l 为训练样本集的样本个数; \mathbf{x}_i 表示输入的分类 特征向量, $\mathbf{x}_i \in \mathbf{x} = \mathbf{R}^n$; $y_i \in y = \{-1,1\}$,为输出的 学习 结 果. 训练 的 目 的 是 寻 找 最 优 判 决 函 数 $f(\mathbf{x}, a)$,满足条件的为具有最大间隔的分类超平面 $w\mathbf{x} + b = 0$,其中 w 为最优分类面的法向量. 若要得 到此超平面,则需求解凸二次规划问题^[15]. 当 2 类样 本不可分时,引入松弛变量 $\boldsymbol{\xi}$,使用拉格朗日因子 法,将优化问题转化成对偶形式

$$\begin{cases} \max L(a) = \sum_{i=1}^{l} a_i - \frac{1}{2} \sum_{i,j=1}^{l} a_i a_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \\ \text{s. t.} \quad 0 \leqslant a_i \leqslant C \qquad i = 1, 2, \cdots, l \\ \sum_{i=1}^{l} a_i y_i = 0 \end{cases}$$
(4)

式中: a_i 和 a_j 为拉格朗日因子;(x_i , y_i)为上述训练 样本集中的一个样本;C为惩罚因子.

如果输入模式在原始空间线性不可分,则通过 引入非线性映射函数 $\phi(x)$ 将低维空间线性不可分 问题映射为高维空间线性可分问题,从而求解上式, 得到非线性分类器

$$f(\boldsymbol{x}) = \operatorname{sgn}\left(\sum_{i=1}^{l} a_i^* y_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b^*\right) \quad (5)$$

式中, a_i^* 为上述优化问题得到的不为零的最优解; b^* 为分类阈值,可以用任一个支持向量求得,或通过 2类中任意1对支持向量取中值求得; $K(\mathbf{x}_i,\mathbf{x}_j) = \phi(\mathbf{x}_i)\phi(\mathbf{x}_j)$ 为满足 Mercer 条件的"核函数".本文选 用径向基核函数得到蓝藻 SVM 分类模型

 $f(\mathbf{x}) = \text{sgn}(\sum_{i=1}^{n} a_i^* y_i \exp(-\gamma || \mathbf{x} - \mathbf{x}_i ||^2) + b^*) \quad (6)$ 参数 γ 和惩罚因子 C 分别预设为 0.4 和 100,其结 果如图 7 所示.

4 分析与结论

经过统计,SVM 模型提取的蓝藻面积为4.818 6 km²,大多发生在湖区域的西南部和北部,这是由于 受风引起的湖水水平循环及夏季盛行的东南风的影 响^[8].比较图 5c 和图 7 可以看出,采用基于 NDI_CB 的 SVM 模型提取的蓝藻面积(图 7)有所减少(减少 的面积为1.58 km²),经过参照假彩色图像和实地探 测数据发现 SVM 模型消除了一些位于湖泊西南部 和北部沿岸误分为蓝藻的水体类别.



图 7 蓝藻识别结果 Fig.7 Extraction result of cyanobacteria bloom

淀山湖蓝藻暴发会导致上海市饮用水危机,对人体健康会造成严重的影响,因此,如何快速、有效地识别蓝藻空间分布是进行蓝藻治理的首要问题.基于以上遥感影像的蓝藻提取实验和分析,得到以下结论:

(1) 当低浓度蓝藻没有完全覆盖水面时,包含的水体信息会导致近红外波段反射率较低,对完全提取蓝藻造成了一定障碍.本文在分析蓝藻和其他 典型地物影像光谱曲线及其特征的基础上,改进传统植被指数,构建了归一化蓝藻指数(NDI_CB).通过 k-均值非监督分类结果可以发现,构建的归一化 蓝藻指数较传统的 NDVI 和 RVI 能更加适用于提取 低浓度蓝藻空间分布信息.

(2) 基于影像光谱特征和 NDI_CB 的 SVM 模型 能够根据小样本训练数据构建更加有效的分类决策 面,既能有效地识别各种密度的蓝藻空间分布信息, 又可以减少非监督分类误分的水体信息.

结果表明,利用遥感影像的处理和分析能够有效地获得较大尺度的特定时间的蓝藻空间分布特征,为湖泊的蓝藻预警和治理提供及时有效的科学依据和技术支持.

进一步的研究需要结合近几年蓝藻暴发严重的 情况,利用多源遥感影像进行暴发前后的时空变化 检测和过程分析并结合气象和实地调查数据进行相 互验证.

致谢 本研究遥感影像由复旦大学生命科学学院屈铭 志同学无偿提供,在此表示衷心感谢!

参考文献:

- [1] 周立国,冯学智,王春红,等.太湖蓝藻水华的 MODIS 卫星监 测[J].湖泊科学,2008,20(2):203.
 ZHOU Liguo, FENG Xuezhi, WANG Chunhong, et al. Monitoring cyanobacteria bloom based on MODIS data in Taihu Lake [J].
 Journal of Lake Science,2008,20(2):203.
- [2] 刘建萍,张玉超,钱新,等.太湖蓝藻水华的遥感监测研究[J]. 环境污染与防治,2009,31(8):79.
 LIU Jianping, ZHANG Yuchao, QIAN Xin, et al. Review of cyanobacteria remote sensing in Taihu Lake[J]. Environmental Pollution & Control,2009,31(8):79.
- [3] 陈云,戴锦芳.基于遥感数据的太湖蓝藻水华信息识别方法
 [J].湖泊科学,2008,20(2):179.
 CHEN Yun, DAI Jinfang. Extraction methods of cyanobacteria bloom in Taihu Lake based on RS data[J]. Journal of Lake Science:2008.20(2):179.
- Vincent R K, Qin X M. Phycocyanin detection from Landsat TM data for mapping cyanobacterial blooms in Erie Lake [J]. Remote Sensing of Environment, 2004, 89:381.
- [5] 谢国清,李蒙,鲁韦坤,等. 滇池蓝藻水华光谱特征、遥感识别及暴发气象条件[J]. 湖泊科学,2010,22(3):327.
 XIE Guoqing, LI Meng, LU Weikun, et al. Spectral features, remote sensing identification and breaking-out meteorological conditions of algal bloom in Dianchi Lake [J]. Journal of Lake Science,2010,22(3):327.
- [6] Aun Reinart, Tiit Kutser. Comparison of different satellite sensors in detecting cyanobacterial bloom events in the Baltic Sea[J]. Remote Sensing of Environment, 2006, 102(1):74.
- [7] Becker Richard H, Sultan Mohamed I, Boyer Gregory L, et al. Mapping cyanobacterial blooms in the Great Lakes using MODIS
 [J]. Journal of Great Lakes Research, 2009, 35(3):447.
- [8] 杨虹,由文辉,汪益嫔,等.淀山湖夏季蓝藻时空分布研究[J]. 安徽农业科学,2010,38(21):11411.

YANG Hong, YOU Wenhui, WANG Yipin, et al. Study on spatial and temporal distribution of blue-green algae community in Lake Dianshan[J]. Journal of Anhui Agriculture Science, 2010, 38(21):11411.

- [9] 马荣华,孔繁翔,段洪涛,等.基于卫星遥感的太湖蓝藻水华时空分布规律认识[J].湖泊科学,2008,20(6):687.
 MA Ronghua, KONG Fanxiang, DUAN Hongtao, et al. Spatiotemporal distribution of cyanobacteria blooms based on satellite imageries in Taihu Lake , China[J]. Journal of Lake Science, 2008,20(6):687.
- [10] 段洪涛,张寿选,张渊智.太湖蓝藻水华遥感监测方法[J].湖 泊科学,2008,20(2):145.
 DUAN Hongtao, ZHANG Shouxuan, ZHANG Yuanzhi.
 Cyanobacteria bloom monitoring with remote sensing in Taihu Lake[J]. Journal of Lake Science,2008,20(2):145.
- [11] Barale V, Jaquet J, Ndiaye M. Algal blooming patterns and anomalies in the Mediterranean Sea as derived from the SeaWiFS data set (1998—2003) [J]. Remote Sensing of Environment, 2008, 112: 3300.
- [12] Guo L. Doing battle with the green monster of Lake Taihu [J]. Science, 2007, 317:1166.
- [13] 彭文祥,王海银,姜庆五.基于多时相卫星数据的太湖蓝藻暴发的动态变化监测[J].复旦大学学报:医学版,2008,35(1):63.
 PENG Wenxiang, WANG Haiyin, JIANG Qingwu. Dynamic change monitoring of cyanobacteria blooms using multi-temporal satellite data in Taihu Lake [J]. Fudan University Journal of Medical Sciences,2008,35(1):63.
- [14] Vapnik N Vladimir . Statistical leaning theory[M]. New York: John Wiley & Sons, 1998.
- [15] 杨朝辉.异源合成孔径雷达景象匹配的评估技术研究[D].上海:同济大学,2009.
 YANG Zhaohui. Research on evaluation technique for multisensor SAR scene matching [D]. Shanghai: Tongji University.2009.