

利用 TM 图像逐级分类提取非点源信息

胡远安¹, 程声通¹, 贾海峰¹, 张海星², 谭斌², 赵江² (1. 清华大学环境科学与工程系环境模拟与污染控制国家重点联合实验室, 北京 100084, E-mail: cst-den@tsinghua.edu.cn; 2. 江西省环境信息中心)

摘要: 探讨了运用逐级分类从 TM 图像中提取与非点源污染有关的土地利用信息的方法。研究表明, 逐级分类与一般的最大似然度分类相比效果更佳, 具有简化复杂问题、突出研究重点和方法灵活等优点, 能够快速有效地提取所需要的信息。

关键词: 非点源; 土地利用; 逐级分类

中图分类号: X87 文献标识码: A 文章编号: 0250-3301(2003)01-05-0152

Extract Non-point Pollution Source Information from a TM Image by Hierarchical Classification

Hu Yuan'an¹, Cheng Shengtong¹, Jia Haifeng¹, Zhang Haixing², Tan Bin², Zhao Jiang² (1. State Key Joint Laboratory of Environmental Simulation and Pollution Control, Department of Environmental Science and Engineering, Tsinghua University, Beijing 100084, China E-mail: cst-den@tsinghua.edu.cn; 2. Jiangxi Environmental Information Center)

Abstract: The issue of the article is hierarchical classification that is used to extract land-use information relating to non-point source pollution from TM image. The result shows that it performs better than traditional maximum likelihood method. It is useful to extract information quickly and effectively with advantages of simplifying complex problem, giving prominence to key point and having more flexibility.

Key words: non-point source; land-use; hierarchical classification

近年来, 流域范畴的非点源污染越来越引起关注。狭义的非点源污染是指在降雨-径流的淋溶和冲刷作用下, 大气、地面和地下的污染物进入江河、湖泊和海洋等水体而造成的污染。非点源污染负荷与土地利用类型密切相关。土地利用类型是自然因素和经济因素的综合体现, 表征了一定时期内相对稳定的地面覆盖物、经济活动、管理措施等与非点源污染密切相关的信息。一般来说, 耕地是贡献率最高的非点源, 其次为城镇、村庄等居民点, 草地、林地产生的污染物负荷随植被覆盖度和植被类型变化。土地利用信息的提取对非点源识别来说至关重要。

遥感是非点源识别空间信息和获取属性信息的重要手段。遥感图像具有覆盖面广、周期快、信息量丰富等特点。遥感技术的迅速发展, 使用户能够获得不同时相、空间分辨率和光谱分辨率的数据, 满足不同时间和空间尺度的模

型要求。另外, 遥感方法以图像像元为单位, 把流域中各项参数(如土地利用、植被分布等)分解到每个像元, 可以很方便地应用于非点源模型计算。20世纪80年代, 遥感技术开始广泛地应用于非点源识别, 并取得了令人满意的成果。我国在利用遥感技术识别水土流失和土壤侵蚀量也有一定的成绩和应用经验^[1]。

本研究的主要目的在于探讨快速准确地获取与非点源污染相关的土地利用信息的方法。

1 研究区概况

研究区域为江西省袁水上游小流域, 多山地丘陵。遥感数据源为 TM (The matic Mapper) 影像, 含 7 个波段数据。成像时间为 1998-10-03, 图幅为东经 113.85 ~ 114.25, 北纬 27.41 ~

基金项目: 国家环境保护总局、江西省环境保护局科研基金项目; 清华大学环境科学与工程研究院 985 基金项目
作者简介: 胡远安(1977~), 女, 博士研究生, 主要从事环境系统分析研究。

收稿日期: 2002-03-08; 修订日期: 2002-05-23

27.71. 成像时间内研究区域的土地利用类型特点如下:主要植被类型为林地,以常绿林为主;水稻为当地的主要作物,正处于收获季节,部分地区已经收割;旱地大多零星分布,面积较小,且作物类型多样,覆盖度变化范围大。

根据非点源污染的特点,在遥感信息提取中,主要关心的内容包括:

- (1) 耕地、居民点——可能的污染负荷最大的土地利用类型。
- (2) 林地——研究区域内最主要的土地利用类型,林地的信息提取包括植被的类型及覆盖度等级等直接影响林地污染负荷的信息。
- (3) 主要的水库——可能对污染物的迁移产生显著的影响。

在本次研究中,对道路、小河流等线性信息的关注程度较低。

2 方法选择

利用 TM 影像光谱信息提取土地利用类型,主要方法有:

- (1) 利用原始波段进行分类,三波段的原

始光谱组合在土地利用分类中应用较多,如 $T_{M_{3-4-5}}$ 和 $T_{M_{2-3-4}}$ 、 $T_{M_{2-4-7}}$ 则适用于城市分类。有部分学者认为 4 波段的组合分类效果更佳, $T_{M_{1-4-5-7}}$ 适用农业用地分类, $T_{M_{2-4-5-7}}$ 适用于森林分类。

- (2) 利用变换后的特征进行分类,主要的变换方法包括主成分分析、缨帽变换、比值变换和生物量提取等等,在林地提取中,生物量提取和比值变换得到了广泛地应用。

利用光谱信息进行分类,其结果的准确度在很大程度上取决于所选取的波段组合和特征提取方法。不同地物类型在不同波段上的可分离性不同,识别不同地物的最佳波段组合和特征提取方法不一致,这直接导致了使用相同的训练区,不同的波段组合可获得的结果不尽相同。在本次研究中,利用多种波段组合对 TM 图像进行了试分类,结果发现,用一种方法一次将有效信息提取出来非常困难,表 1 是其中一次试分类所得到的训练区混淆矩阵。

由表 1 可知,在对林地 1 的划分中出现了

表 1 最大似然度分类混淆矩阵/ %

Table 1 Contingency matrix of maximum likelihood classification/ %

样本	居民点	水体	耕地	裸地	林地 1	林地 2	林地 3
居民点	99.50	0.00	0.00	0.00	0.00	0.00	0.00
水体	0.50	100.00	0.00	0.00	30.25	0.00	0.00
耕地	0.00	0.00	99.13	0.00	0.00	0.00	0.00
裸地	0.00	0.00	0.00	100.00	0.00	0.00	0.00
林地 1	0.00	0.00	0.00	0.00	54.53	0.00	0.00
林地 2	0.00	0.00	0.87	0.00	15.22	99.07	2.07
林地 3	0.00	0.00	0.00	0.00	0.00	0.93	97.93

较大的误差,有相当一部分被误分为水体。

为解决以上问题,在本次研究中采用逐级分层分类提取,逐级分类的优点在于可以充分利用光谱信息,把复杂的问题划分为相对简单的问题,针对不同的分类目标选择最佳的波段组合,避免一次划分多种类别在选择波段或特征参数时遇到的困难^[2]。该方法在研究区域土地利用类型较为复杂,光谱信息(遥感数据源)较丰富的情况下常用。

3 逐级分类

本次研究中的分类步骤见图 1。

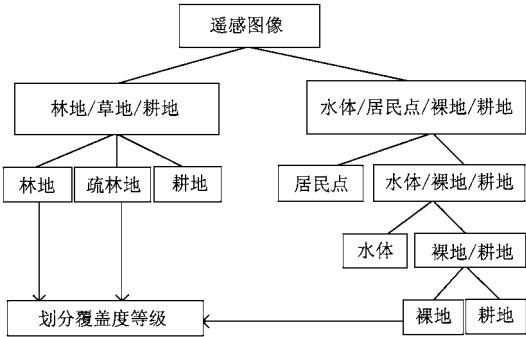


图 1 逐级分类流程

Fig.1 Flow of hierarchical classification

3.1 植被与非植被的分离

利用植被指数可以有效地分离植被与非植被.由图 2 可以看出,归一化植被指数(Normal-

ized Difference Vegetation Index ,NDVI) 与植被的茂密程度呈对应关系. 在本次研究中,尝试了 2 种方法.一是利用

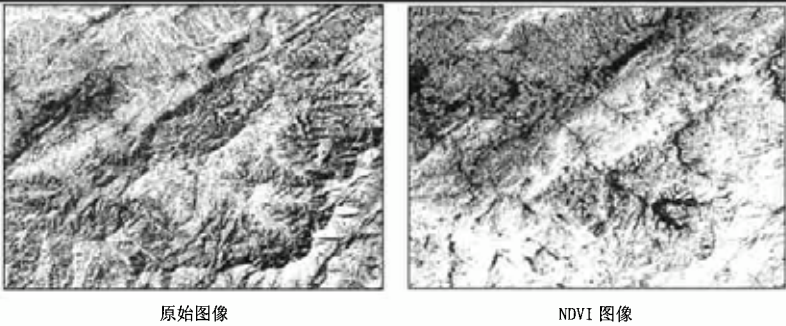


图 2 研究区域原始图像与 NDVI 图像
Fig.2 Raw image and NDVI image of study region

阈值法,根据 NDVI 值直接划分植被与非植被,在本次研究中选取的阈值为 0.14,这是已经收割的耕地在图像上的 NDVI 值;二是利用 NDVI, TM_4/TM_3 , TM_5/TM_3 进行非监督分类,

将非监督分类结果与原图进行比较,在本次研究中,将遥感图像分为 10 类.提取结果见图 3. 分离结果 1 中黑色为非植被,白色为植被;分离结果 2 中,黑~白分别代表 1~10 类.

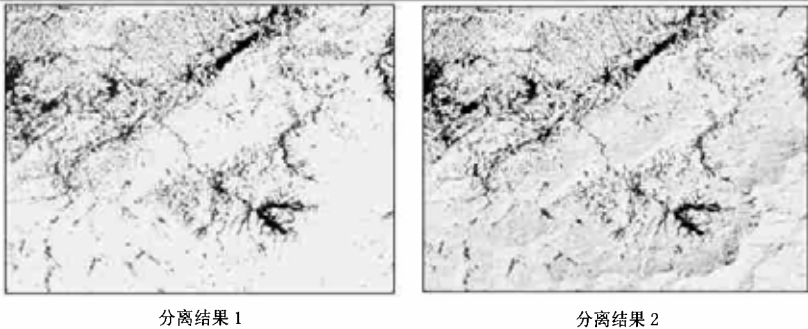


图 3 不同方法的分离结果
Fig.3 Results of different methods

经对比发现,第 1 类为非植被,其它皆为植被.对比以上 2 种方法划分的结果,基本一致.第 2 种方法对道路等线性信息的提取优于前者.由于在本次研究中,对道路的提取要求很低,因此可认为这 2 种方法等价. 对比原遥感影像可以看出,由于研究区域多山地丘陵,部分阴影的影响难以根据 NDVI 去除.在工作中,必须注意对阴影部分的处理.在本次研究中,根据阴影部分总体灰度低的特点,成功地提取了阴影部分,对阴影部分的提取并不生成某一土地利用类型,而是留待后续处

理.由于阴影部分的信息往往存在着较大的失真,在没有类似的训练区的情况下,需要根据周围地物和其它信息综合判断. 3.2 非植被部分的进一步分类 非植被部分主要包括居民点、水体和裸地.其典型波谱如图 4 所示.由图中可以看出,3 类地物存在着较明显的差异,裸地几乎每一波段的灰度值都大于其它地物,居民点的灰度值居中,水体最低. (1)居民点提取 对于 TM 图像,居民地大多为混合像元,实际上是各类建筑物及周围的

道路、耕地、水体和林地等物的综合表现.根据文献[3]的研究,利用居民地具有 TM_2 、 TM_4 、 TM_7 非常接近的谱间关系特征,可以用最简单的阈值法提取居民点信息,构建表达式如下:

$$\begin{aligned} TM_2 - TM_4 &< T_1 \\ TM_4 - TM_7 &< T_2 \\ TM_7 - TM_3 &< T_3 \end{aligned}$$

在本次研究中,同样利用居民点的上述特征,构建如下表达式:

$$\sum_{i=2,4,7} \left| TM_i - (TM_2 + TM_4 + TM_7) / 3 \right| < T$$

选择 T 值为 20.利用以上方法基本上能够成功地提取居民点信息.

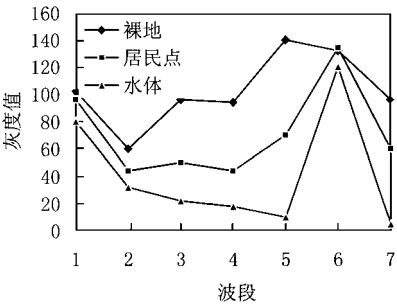


图 4 非植被部分主要地物光谱
Fig.4 Spectrum of non-vegetation

(2) 水体提取 本研究中,遥感识别对河流的提取要求较低.一般来说,只要求提取对污染物截留作用较大的湖泊、池塘等,河流网系主要依赖 DEM 生成.

对去除居民点后的非植被部分,根据水体 TM_5 波段灰度值远低于其它地物的特征,利用阈值法提取水体.本次研究中选取的阈值为 35.从分类结果看,能够成功提取流域内的湖泊及面积较大的池塘,满足要求.

(3) 裸露地与耕地提取 利用监督分类,将扣除居民点及水体后的非植被部分进一步分为 2 类:已经收割的耕地;植被覆盖度极低的裸露地,一般为复杂的混合像元,往往包括了道路、未利用地、建筑、疏林地等多种地物.

3.3 植被部分的进一步分类

由于成像期间部分水稻还未收割,因此这一部分耕地的信息是水稻的光谱.植被部分的

典型地物光谱如图 5 所示.可以发现,植被之间的光谱差异较非植被小,其光谱差异主要集中在 3、4、5 波段.同时由于日照与阴影的影响,存在着较为显著的同物异谱现象.在本次研究中根据比值变换后的结果分类,利用 TM_3/TM_2 , TM_4/TM_3 , TM_5/TM_4 , TM_7/TM_5 进行最大似然度监督分类,分离耕地、林地及疏林地.试验区的混淆矩阵如表 2 所示,与表 1 相比,分类精度明显改善.

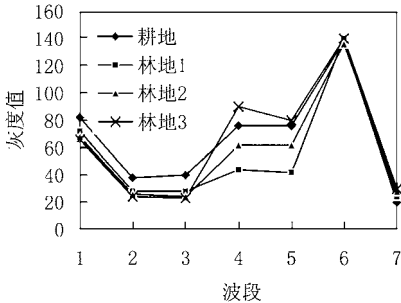


图 5 植被部分地物光谱
Fig.5 Spectrum of vegetation

表 2 比值变换混淆矩阵/%

Table 2 Contingency matrix of classification using ratios/%				
样本	耕地	林地 1	林地 2	林地 3
耕地	99.51	0.00	0.00	0.00
裸地	0.25	0.00	0.00	0.00
林地 1	0.12	94.56	0.19	0.00
林地 2	0.12	0.53	99.61	2.18
林地 3	0.00	4.90	0.19	97.82

3.4 覆盖度划分

植被指数与覆盖度之间存在着很强的相关性,可以根据植被指数来划分覆盖度等级^[4].常见的植被指数除 NDVI 外,还有土壤修正植被指数^[5](Soil Adjusted Vegetation Index, SAVI),其表达式如下:

$$\begin{aligned} NDVI &= (TM_4 - TM_3) / (TM_4 + TM_3) \\ SAVI &= (TM_4 - TM_3)(1 + L) / \\ &\quad (TM_4 + TM_3 + L) \end{aligned}$$

其中, L 为土壤调节参数,建议值为 0.5.根据文献[6]的研究,NDVI 在图像大部分区域覆盖度较低时提取效果最佳;当图像大部分区域为

高覆盖度植被且变化范围较小时,采用 L 为 0.5 的 SAVI 可以获得满意的效果;对于除上述 2 类情况以外的大多数 TM 图像而言,采用 L 为 5 的 SAVI 可以满足要求.

从上述分类结果中提取出裸地、疏林地、林地的信息,在相应的 SAVI($L = 5$) 上进行非监督分类.对比遥感图像,确定每一类所对应的覆盖度等级.其对应的植被指数范围如表 3.

表 3 植被覆盖度对应的植被指数

Table 3 Corresponding vegetation indices to different green cover					
覆盖度类型	裸露地	极稀疏植被	稀疏植被	中密植被	密植被
SAVI	- 0.4 ~ 0.8	0.8 ~ 1.3	1.3 ~ 1.8	1.8 ~ 2.4	2.4 ~ 3.7
NDVI	- 0.04 ~ 0.15	0.13 ~ 0.26	0.25 ~ 0.35	0.33 ~ 0.44	0.42 ~ 0.61

3.5 分类结果叠加

将各级分层提取的信息进行叠加,并根据知

识判断阴影部分的土地利用类型,就可以得到最后的分类结果图.各部分地物的比例如表 4.

表 4 土地利用分类结果

Table 4 Result of land use classification								
地物类型	居民点	水体	耕地	裸露地	极疏植被	稀疏植被	中密植被	密植被
所占面积/ %	1.9	0.7	22.2	1.3	6.3	20.2	27.7	19.7

4 结论

非点源的遥感识别重点有别于一般的土地利用类型识别,它对地物的类型划分与非点源污染特点和研究目标密切相关.

分级提取地物信息更加有利于突出重点,由于不同地物对非点源污染负荷的贡献率不一,因此在遥感图像分类时,对于不同地物的识别精度要求也不一样.分级提取有助于保证重点地物提取的精度.

另外,分级提取信息有利于将复杂的问题简单化,可以针对不同的分类目标选择不同的分类方法.从本研究来看,在分层提取的过程中,选用阈值法和最大似然度法已经基本上可以成功地分离所需地物类型.

最后,分层提取也有利于对分类结果的灵

活处理,对地物的合并和进一步划分较简便,方便用户根据研究区域的特点和模型需求划分所需的地物类型.

参考文献:

1 卜兆宏,孙金庄,董勤瑞等.应用水土流失定量遥感方法监测山东全省山丘区的研究.土壤学报,1999,36(1): 1 ~ 8.

2 王长耀,牛铮,唐华俊等.对地观测技术与精细农业.北京: 科学出版社,2001.82 ~ 100.

3 杨存建,周成虎.TM 影像的居民地信息提取方法研究.遥感学报,2000,14(2): 146 ~ 150.

4 Colwell J E. Vegetation canopy reflectance. Remote Sens. Environ., 1974,3: 175 ~ 183.

5 Huete A R. A soil adjusted vegetation index(SAVI). Remote Sens. Environ., 1988,25(2): 295 ~ 309.

6 高志海,魏怀东,丁峰.TM 影像 VI 提取植被信息技术研究.干旱区资源与环境,1998,12(3): 98 ~ 104.