

# 基于空间概率面的山区居民地遥感信息提取

张源, 王仰麟, 彭建, 吴健生

(北京大学环境学院, 北京 100871; 北京大学深圳研究生院数字城市与景观生态中心, 广东 深圳 518055)

**摘要:** 根据地物之间光谱特征建立的基于知识的遥感居民地信息提取模型是目前居民地信息提取中最普遍的方法, 但由于高程差异的影响, 其在山区居民地信息提取中效果不理想。以云南省丽江市部分地区为例, 在 GIS 支持下, 通过构建多因素空间概率面的方式, 综合运用地形和光谱特征信息实现山区居民地遥感信息提取。结果表明, 地形差异是影响山区居民地信息提取精度的最主要因素, 其影响程度占有所有影响因素的 50% 强; 在光谱信息识别的基础上, 引入地形这一辅助信息, 运用空间概率面能够有效地改善山区居民地信息的提取效果, 识别精度从 57.5% 提高到 82.5%。

**关键词:** 空间概率面; 山区居民地; 遥感; 地理信息系统

**中图分类号:** TP79; P208      **文献标识码:** A      **文章编号:** 1672-0504(2006)04-0006-05

## 0 引言

居民地是人类从事生产和生活需要而集聚定居的各种形式的居住场所<sup>[1]</sup>, 是自然景观和人造景观的综合体。及时、准确地掌握居民地的空间分布信息可以为灾害评估、城镇扩展和环境变化等相关研究提供必要的基础信息<sup>[2-6]</sup>, 因此利用遥感影像快速准确地提取居民地信息具有重要意义。一般 NOAA 影像上能识别县城以上级的居民地, Landsat TM 影像能识别乡镇级的居民地, 雷达影像上则可识别具有 1 000 人以上的居民地<sup>[7]</sup>。

根据不同类型地物的影像光谱特征建立的基于知识的遥感居民地信息提取模型是目前居民地信息提取中最普遍的方法。在高差变化不显著的区域(如平原地区), 其提取效果较理想<sup>[8-13]</sup>; 但在高差变化显著的区域(如山区), 其提取效果则不尽如人意<sup>[14]</sup>。这是因为在高差变化显著的区域, 居民地光谱特征除受自身内部组成成分<sup>[8]</sup>和影像质量<sup>[9]</sup>影响外, 还受地形因素的显著影响, 即由坡度变化和太阳高度角所带来的辐射差异, 使得山区居民地光谱信息易于同其他地物混淆, 从而降低了其光谱可分性。因此, 山区居民地遥感信息提取一直是遥感专题信息提取的难点之一, 针对这一问题的研究尚不多见。另一方面, 由于遥感影像光谱信息不完备而引入其他辅助信息进行相关研究, 一直是遥感信息解译中的常用方式<sup>[15-18]</sup>。

本文以地处横断山区的云南省丽江市部分地区

为试验区, 探讨在 GIS 支持下, 通过构建多因素空间概率面的方式, 综合运用地形和光谱特征信息, 实现山区居民地遥感信息的提取。

## 1 研究方法与数据

### 1.1 空间概率面

空间概率是针对某一类信息而言, 指在试验区内的空间上该类信息在某一空间单元或某一点上存在的概率。而空间概率面是指整个试验区所有空间单元的空间概率所构成的一个空间上连续的面。其形式上与 DEM 一致, 只是在空间上表示的不是高程值, 而是针对某一类信息存在与否的概率值。本研究针对居民地信息提取, 重要的是确定居民地在试验区内的空间概率面的状况。

一般某类信息实际的空间概率面的形状通常受多因素综合影响; 对于每种因素, 都存在一个空间概率面。涉及这种多因素概率综合, 即数学上所谓的不确定性推理问题, 通常依据的是贝叶斯定理 (Bayes' s theorem), 可以简略为如下表达式:

$$P(H_i | E_j) = \frac{P(E_j | H_i) \times P(H_i)}{P(E_j)} \quad (1)$$

$$P(E_j) = \sum_{i=1}^n P(E_j | H_i) \times P(H_i) \quad (2)$$

式中:  $H$  为假设;  $E$  为证据;  $P(H_i)$  为某假设  $i$  的概率 (probability);  $P(E_j)$  为某证据  $j$  的概率;  $P(E_j | H_i)$  为某证据对于给定假设的概率, 又称条件概率 (conditional probability);  $P(H_i | E_j)$  为一个假设在给定证据条件下的概率。

收稿日期: 2006-05-08; 修订日期: 2006-06-21

基金项目: 国家“973”重点基础研究规划项目(G2000046807); 国家自然科学基金项目(40471002)

作者简介: 张源(1980-), 男, 硕士研究生, 主要从事 3S 技术在城市规划和景观生态学中的应用研究。E-mail: yzhang\_geo@gmail.com

本研究针对居民地信息提取, 考虑的影响因素是光谱和地形信息, 因此式(1)可以改写为:

$$P_i = f(RS, terrain_i) = \frac{P_{rsi} * P_{terrain_i}}{\sum_i P_{rsi} * P_{terrain_i}} \quad (3)$$

式中:  $P_i$  是空间上  $i$  点为居民地的空间概率;  $P_{rsi}$  是基于光谱信息得到的该居民地的空间概率;  $P_{terrain_i}$  是基于地形信息的该居民地的空间概率。

在实际研究中, 地形信息主要选取高程和坡度数据, 因此, 式(3)又可以改写为:

$$P_i = f(RS, alti, slope) = \frac{P_{rsi} * P_{altii} * P_{slopei}}{\sum_i P_{rsi} * P_{altii} * P_{slopei}} \quad (4)$$

式中:  $P_{altii}$  和  $P_{slopei}$  分别是基于高程和坡度的居民地的空间概率。

### 1.2 阈值的确定

空间概率面反映的是某类信息在空间上存在的可能性的状况, 仅是概率而已, 无法确定该类信息在空间上的实际分布状况。而确定信息的实际分布状况, 则意味着根据特定的概率阈值将原来连续的空间概率面转化为简单的 0 或 1 的二值面。因此, 阈值的确定是信息提取的关键。

阈值的确定一般有两种方法: 1) 临界概率限值法, 即不考虑总量的限制, 只要空间概率达到一定的阈限值即确定该信息是否存在; 2) 总量控制法, 即认为某类信息在特定时间内其总量是确定的, 根据总量要求, 按照每个空间单元的空间概率值确定实际分布状况。本研究采用第 2 种方法, 显然, 在特定的时间范围内, 试验区内居民地的总面积是一定的。

### 1.3 数据来源与处理

试验区位于丽江市东北部(图 1), 地处云贵高原西缘, 由北部东北—西南走向的玉龙山脉、西部西北—东南走向的云岭山脉和山脉间的丽江坝子构成。区内地貌类型多样, 高程差异显著, 最高点玉龙雪山主峰海拔 5 596 m, 最低点金江乡席草地附近金沙江



图 1 试验区示意  
Fig 1 Study area

水面海拔 1 219 m, 相对高差 4 377 m; 该区地势险峻, 最大坡度为 78.3°。区内土地覆被类型包括河流、河漫滩、湖泊、草地、林地、积雪与冰川、高山草甸、流石滩、水田、旱地和居民地等。

本研究使用的基础数据为 1996 年 2 月 16 日的 Landsat 5 遥感图像, 轨道号 131/41(图 2), 分辨率为 30 m, 选取 1~5、7 波段进行研究; 以及 1994 年 1:10 万丽江市土地利用图、1996 年丽江市 1:25 万国家基础地理数据库。

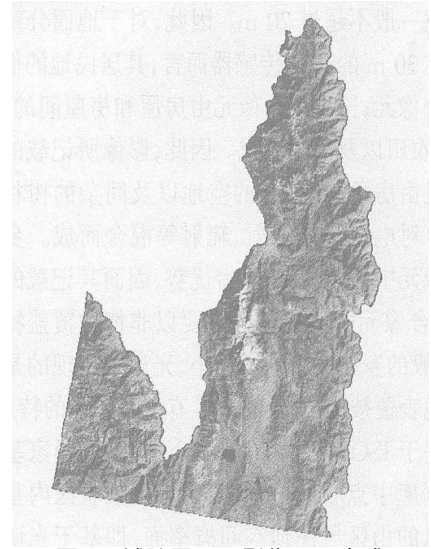


图 2 试验区 TM 影像 742 合成  
Fig. 2 TM image (band 742) in the study area

研究流程为: 1) 基于 Landsat TM 影像光谱信息得到试验区内居民地空间分布二值图; 2) 从 1:25 万国家基础地理信息数据库中选出试验区内点状居民地信息和等高线信息, 计算综合高程和坡度信息与居民地之间的概率分布, 从而构建山区居民地在试验区内的空间概率面; 3) 根据统计数据中试验区内居民地总面积确定概率阈值, 根据阈值将总面积分配到整个试验区范围内, 最终得到山区居民地的空间分布信息。

需要指出的是, 由于试验区内土地覆被类型复杂且居民地面积较小, 加上高程差异所带来的“同谱异物”效应显著, 使得基于人工训练区选择的监督分类并不适合山区居民地信息提取。因此, 本研究中采用非监督分类的 ISODATA 算法进行预分类。为避免漏分, 在预分类时结合点状居民地空间位置信息。

## 2 结果及分析

### 2.1 基于光谱信息的空间概率面

试验区主要有城镇和乡村两类居民地, 城镇包

括丽江城区和各个乡镇, 乡村则主要是行政村和自然村。丽江城区分为新城和古城两部分。新城地表覆被主要由楼房、水泥地面、裸土地、草坪和零星树木组成, 楼房主要是水泥顶并铺有沥青。而丽江古城与一般的乡村居民地类似, 主要由房屋、空地及零星树木组成, 房屋以平房为主, 房顶多为青瓦。路两边的居民地多为并排分布, 其他的为随机、散乱分布。

无论是乡镇还是乡村地区, 房屋与房屋之间的空地宽一般不超过 20 m。因此, 对于地面分辨率为 30 m×30 m 的 TM 传感器而言, 其居民地的像元多为混合像元, 这种混合像元由房屋和房屋间的空地、绿地、农田以及林木组成。因此, 影像所记载的遥感信息是由房顶、房屋间的空地以及间杂的树林、草、农田等对电磁波的反射、辐射等混合而成。乡镇居民地像元中房顶、空地等占优势, 因而其记载的光谱中, 混合像元的光谱特性主要以非植被覆盖特点为主; 一般的乡村房屋规模较小, 光谱中出现的是以上多种地表覆被的混合, 并多具有植被光谱的特点。

基于 ISODATA 算法, 结合 1:25 万国家基础地理数据库中点状居民地信息, 得到试验区内基于光谱信息的山区居民地空间概率面, 即基于光谱特征信息的居民地分类结果(图 3)。需要指出的是, 这里的空间概率事实上只有 0 和 100% 两种情况, 而这种空间分布的二值图可以认为是一种特殊情况下的空间概率面。

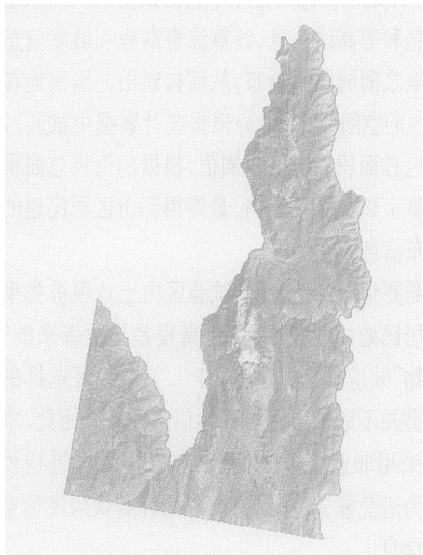


图 3 基于光谱的居民地信息提取  
Fig. 3 Residential area extraction based on spectral analysis

### 2.2 基于地形信息的空间概率面

将 1:25 万国家基础地理数据库中点状居民地信息与 DEM 进行叠加分析, 得到不同高程居民地分

布概率(图 4、图 5)。试验区高程范围为 1 300 ~ 5 400 m, 居民地分布在 1 300 ~ 3 600 m 高程范围内。其中有两个峰值: 高程 1 800 ~ 2 000 m, 16.3% 的居民地分布在该范围; 高程 2 300 ~ 2 500 m, 25.5% 的居民地集中分布在该范围。通过 GIS 空间信息提取发现, 1 800 ~ 2 000 m 范围内的居民地都集中在试验区内两条河流两岸的谷地和阶地上, 而 2 300 ~ 2 500 m 范围内的居民地则集中在试验区内的山谷平坝中。同时, 在高程超过 2 500 m 的区域, 居民地分布数量随高程增加而减少, 在 3 400 m 以上居民地分布数量小于 1%。

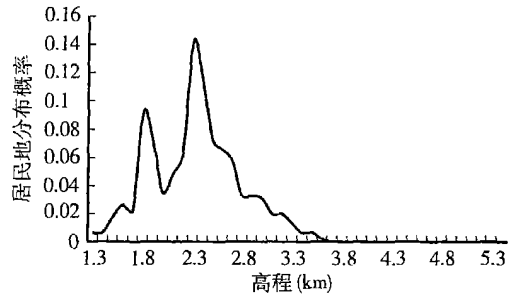


图 4 不同高程居民地分布概率  
Fig. 4 Frequency of distribution of residential area in different altitude

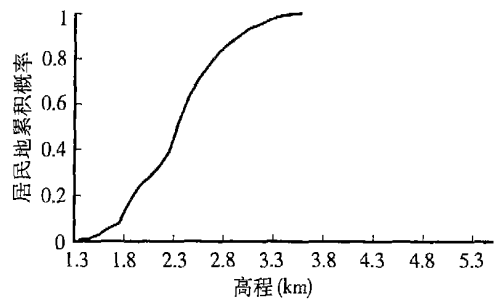


图 5 不同高程居民地分布累积概率  
Fig. 5 Quantile of distribution of residential area in different altitude

将点状居民地信息再与坡度分布图叠加, 可以得到不同坡度居民地分布概率(图 6、图 7)。试验区内居民地仅分布在 0° ~ 40° 的坡度范围内, 超过 95% 的居民地分布在坡度小于 30° 的地区。其中 28.7% 的居民地分布在平地, 坡度大于 15° 地区随坡度增加居民地逐渐减少。

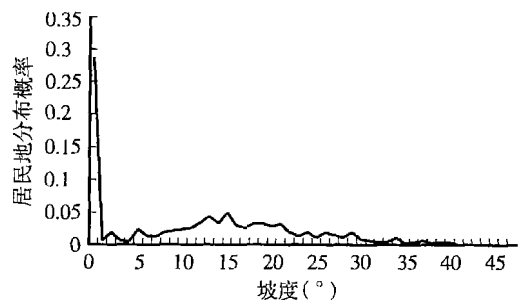


图 6 不同坡度居民地分布概率  
Fig. 6 Frequency of distribution of residential area in different slope

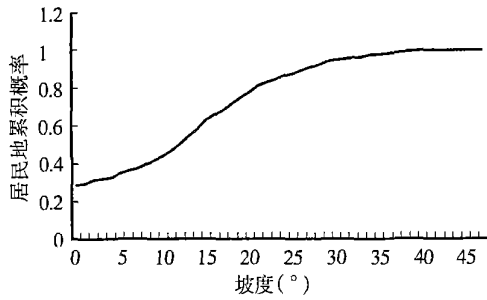


图7 不同坡度居民地分布累积概率

Fig. 7 Quantile of distribution of residential area in different slope

### 2.3 结果检验

根据统计数据, 试验区内居民地总面积为  $6\,606.55\text{ hm}^2$ , 从而确定阈值为  $64.1\%$ , 以此最终确定居民地的空间分布信息。将基于光谱特征信息的分类结果(图3)和通过空间概率面得到的结果(图8), 与1994年1:10万丽江市土地利用图对比, 进行精度检验。在空间生成250个随机点, 其中150个点位于1:10万土地利用图上居民地所在区域, 其余100个点位于非居民地区域。结果显示, 仅依靠光谱进行预分类的精度为  $57.5\%$ , 而通过地形信息修正的空间概率面得到的结果分类精度达  $82.5\%$ , 大大提高了山区居民地信息提取的精度。进一步分析发现影响试验区内山区居民地信息提取精度的因素主要有阴影、休耕地、云雾和河漫滩。

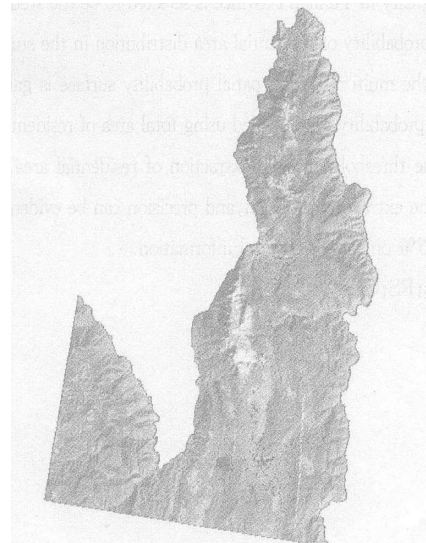


图8 基于空间概率面的居民地信息提取

Fig. 8 Residential area extraction based on spatial probability surface

如表1所示, 阴影是影响山区居民地信息提取精度的最主要因素, 影响程度达  $50\%$  强, 这是由高程差异显著造成的; 通过空间概率面提取后, 其影响减少了近  $90\%$ 。对于其他3种影响因素, 可以通过选择遥感影像时相避免休耕地的影响, 通过成像时的天气状况修正云雾的影响, 通过与河流的位置关系修正河漫滩的影响, 但这些都不是本研究中关注的

问题。

表1 影响分类精度的主要因素  
Table 1 Major influence on extraction precision

影响因素	阴影	休耕地	云雾	河漫滩
预分类(%)	22.5	5	7.5	7.5
空间概率面(%)	2.5	5	5	5

### 3 结论

研究表明, 从遥感影像中提取山区居民地信息, 地形的复杂性是影响提取精度的最重要因素。仅依靠光谱信息进行分类, 达不到理想的分类精度; 本研究采用的空间概率面综合考虑了光谱信息和地形要素的影响, 大大提高了提取精度, 用于山区居民地信息提取是可行的。

地理学不仅研究总量的变化, 更重要的是这种变化在空间上的分布。对于总量的变化, 可以通过分析统计资料、生成参数化回归模型或构建非参数化模型(如系统动力学模型)等方法进行研究; 而空间分布研究一直是地理学研究中的热点和难点。本研究采用空间概率面, 是将实际的地理空间根据研究需要视为一个空间概率面, 然后通过总量控制来确定概率阈值, 从而得到山区居民地信息在空间上的分布; 而实际上, 空间概率面比确定的空间分布图更能揭示研究对象在空间上的分布特征。

### 参考文献:

- [1] 左大康. 地理学词典[M]. 北京: 科学出版社, 1990.
- [2] 徐雪仁, 万庆. 洪泛平原农村居民地空间分布特征定量研究及应用探讨[J]. 地理研究, 1997, 16(3): 47-53.
- [3] 田光进, 刘纪远, 张增祥, 等. 基于遥感与GIS的中国农村居民点规模分布特征[J]. 遥感学报, 2002, 6(4): 307-312.
- [4] 房世波, 潘剑君, 陈彩虹. 利用TM和SPOT遥感影像对南京市城镇用地扩展的监测[J]. 南京农业大学学报, 2000, 23(3): 49-52.
- [5] YEH A G O, LI X. An integrated remote sensing and GIS approach in the monitoring and evaluation of rapid urban growth for sustainable development in the Pearl River Delta, China[J]. International Planning Studies, 1997, 2(2): 193-210.
- [6] 戴昌达, 唐伶俐, 陈刚, 等. 卫星遥感监测城市扩展与环境变化的研究[J]. 环境遥感, 1995, 10(1): 1-7.
- [7] HENDERSON F M, XIA Z G. SAR application in human settlement detection, population estimation and urban land use pattern analysis: A status report[J]. IEEE Geoscience and Remote Sensing, 1997, 35(1): 79-85.
- [8] 赵萍, 冯学智, 林广发. SPOT卫星影像居民地信息自动提取的决策树方法研究[J]. 遥感学报, 2003, 7(4): 309-315.
- [9] 杨存建, 周成虎. 基于知识发现的TM图像居民地自动提取研

- 究[J]. 遥感技术与应用, 2001, 16(1): 1—6.
- [10] 高永光, 祝民强, 朱骥, 等. 赣中红壤区 TM 图像的居民地信息自动提取专家模式研究[J]. 国土资源遥感, 2002, 54(4): 67—69.
- [11] 杨存建, 周成虎. TM 影像的居民地信息提取方法研究[J]. 遥感学报, 2000, 4(2): 146—150.
- [12] 陈敏, 刘秉翰, 杨靛青. TM 遥感影像中居民点的自动提取[J]. 福州大学学报(自然科学版), 2004, 32(增刊): 95—98.
- [13] 安如, 赵萍, 王慧麟, 等. 遥感影像中居民地信息的自动提取与制图[J]. 地理科学, 2005, 25(1): 74—80.
- [14] 胡文英, 角缘梅. 基于 ETM+ 土地利用与土地覆被遥感信息提取研究[J]. 云南地理环境研究, 2005, 17(6): 34—37.
- [15] 曹五丰, 秦其明. 基于知识的卫星数字图像公路信息提取研究[J]. 北京大学学报(自然科学版), 1998, 34(2—3): 254—263.
- [16] KONTOES C, WILKINSON G G, BURRILL A, et al. An experimental system for the integration of GIS detain knowledge—based analysis for remote sensing of agriculture[J]. International Journal of Geographical Information Science, 1993, 7(3): 247—262.
- [17] WILKINSON G G, MEGIER J. Evidential reasoning pixel classification hierarchy—apotential method for integrating image classifiers and expert system rules based on geographic context[J]. International Journal of Remote Sensing, 1990, 11(3): 1963—1968.
- [18] FRANKLIN S E, PEDDLE D R. Spectral texture for improve class discrimination in complex terrain[J]. International Journal of Remote Sensing, 1989, 10(8): 1438—1443.

### Research on Extraction of Residential Area in Mountainous Areas Using Spatial Probability Surface

ZHANG Yuan, WANG Yang—lin, PENG Jian, WU Jian—sheng

(College of Environmental Sciences, Peking University, Beijing 100871;

Center of Digital City & Landscape Ecology, Peking University, Shenzhen 518055, China)

**Abstract** Rapidly and accurately acquiring distribution of residential area is very important for lots of researches such as disaster evaluation, urban expansion and environmental change. The development of remote sensing technology provides a rapid and low—cost way to identify and extract residential areas. Nowadays extraction of residential areas from remote sensing images is mainly based on spectral analysis. This methodology is quite effective in areas with little differences in altitude such as plains while it is lost when dealing with areas which distinctly differ in different altitude, such as mountainous areas because the radiation quantities are observably affected by slope and solar altitudinal angle. Therefore other secondary data needs to be involved to improve precision when identifying residential area in mountainous areas. In this paper, mountainous area of Lijiang City in Yunnan Province is selected to be the study area. Based on GIS, three kinds of spatial probability surfaces concerning spatial probability of residential area distribution in the study area are calculated by spectral information, altitude and slope respectively. And the multi—factors spatial probability surface is gained by those three different ones according to Bayes's theorem. Then threshold of probability is confirmed using total area of residential area in the study area, and the total area in space is distributed according to the threshold namely, extraction of residential area. Result shows that vast differences in altitude is the most prominent influence on the extraction precision, and precision can be evidently improved using special probability surface, which is 82.5% and higher to 57.5% only using spectral information.

**Key words:** spatial probability surface; residential area in mountainous areas; RS; GIS