

基于 PCA/NDVI 的森林覆盖 遥感信息提取方法研究

邓书斌¹, 武红敢², 江涛¹

(1. 山东科技大学, 青岛 266510; 2. 中国林业科学院资源信息研究所, 北京 100091)

摘要: 以大兴安岭为试验区, 提出将主成分分析(PCA)得到的第1分量、归一化植被指数(NDVI)和 Landsat TM 1 ~ TM 7 某一波段进行合成, 增强森林覆盖区和背景区信息的反差, 并利用最大似然法对影像进行监督分类, 分类精度超过 92%。通过对不同云雾量和森林覆盖的 2 个时相影像试验表明, 本方法提高了遥感影像森林覆盖信息提取的自动化程度和精度。

关键词: 主成分分析; NDVI; 波段组合; 监督分类; 邻域分析; 最大似然法

中图分类号: TP 79; S 771.8 **文献标识码:** A **文章编号:** 1001-070X(2007)02-0082-04

0 引言

森林是陆地生态系统的主体, 是人类重要的环境资源和物质资源, 随着人类活动范围的扩大和自然灾害的频繁发生, 地表森林遭到严重破坏。为了保护逐年减少的森林资源, 首先要做的工作是掌握森林资源信息, 而传统的地面调查已经跟不上森林变化速度。森林常是遥感观测和记录的第一层面信息, 也是遥感图像反映的最直接信息之一。在多时相、多波段遥感影像上提取森林信息, 具有现势性强、更新周期短和成本相对低廉的特点。传统监督和非监督分类方法精度较低, 而新的分类方法, 如人工神经网络和专家系统, 由于较为复杂, 可操作性较差, 目前还处于方法试验阶段。为此, 本文寻求一种精度更高、步骤更简便的森林信息提取方法, 用于提高遥感影像的森林覆盖信息提取的自动化程度和精度。

1 数据源与试验区

1.1 数据源

为了便于比较, 本研究采用 2 景同季相遥感数据, 分别是 2000 年 6 月 15 日的 Landsat-5 图像和 2005 年 6 月 2 日的 TM 图像, 轨道号均为 WRS120-23。选取 2 景图像中 2518 像元 \times 1589 像元的子

区为研究实验区, 其中心位于北纬 $51^{\circ}58'$, 东经 $125^{\circ}56'$, 面积约 2 480 km²。研究区内 2000 年数据有少量的薄云雾(占研究区总面积比重少于 20%), 大部分区域被森林覆盖; 2005 年数据无云, 其中 40% 以上区域没有森林覆盖。因此, 所选取的 2 个时相同地域试验区差异较大, 以便验证本方法在以下 3 个条件下的可行性:

- (1) 试验区有少量的薄云雾(少于 20%);
- (2) 森林覆盖面积大, 在 80% 以上;
- (3) 森林覆盖面积较小, 在 50% 以下。

1.2 试验区概况

我国东北林区位于欧亚大陆东北部的高纬度地区, 冬季漫长, 夏季气候温热, 雨量充沛。春秋两季较短, 全年无霜期由南向北递减, 全林区有效积温在 $1\ 600^{\circ}\text{C}$ 左右, 年均降水量 400 ~ 600 mm。这些自然条件都有利于树木和林下植物生长。东北林区地带性植被属于寒温带针叶林和温带阔叶混交林, 拥有全国面积最大的天然林。本文研究区地处黑龙江省大兴安岭林区的韩家园林业局部分林场。

2 研究方法

森林覆盖信息提取方法主要包括数据预处理、PCA/NDVI 运算、波段组合、监督分类、精度评价和分类后处理等。其研究方法流程如图 1 所示。

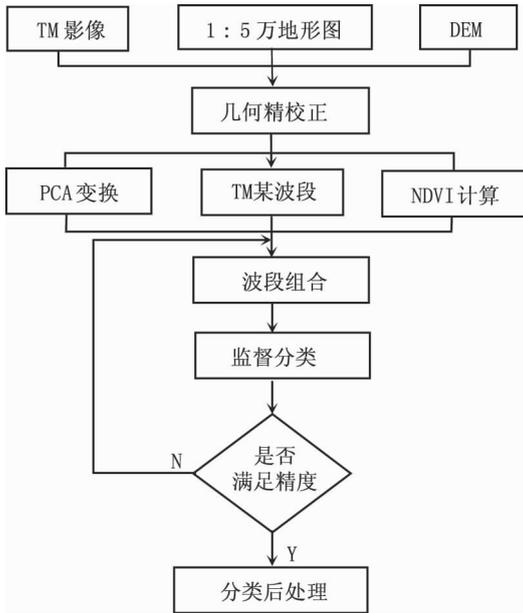


图 1 研究方法流程

2.1 主成分分析(PCA)

TM 影像有 7 个波段,但某些波段之间相关性很高,造成信息冗余。主成分分析(PCA, Principal Components Analysis)又称 KL(Karhunen - Loeve)变换,可以在总信息量不变的前提下,通过正交变换,使各个波段信息重新分配,将多波段的图像信息压缩到比原波段更少的几个分量上。实现在尽可能不丢失信息的同时,用几个综合性分量代表多波段的原图像。试验中,对已经经过辐射校正、几何校正的 2 期 TM 影像分别进行主成分分析(TM 6 是热红外波段,分辨率低不予考虑),得到 6 个分量,其中前 3 个分量集中了影像大部分信息(>95%),第 1 分量信息量最大(>80%)。

2.2 归一化植被指数(NDVI)

遥感图像上的森林信息,主要通过绿色植物叶子和植被冠层的光谱特性及其差异和变化而反映的,研究森林信息,常常选用多光谱遥感数据经分析运算(加、减、乘、除等线性或非线性组合方式),产生某些对森林长势、生物量等有一定指示意义的数值——植被指数。在植被指数中,通常选用的是对

绿色植物强吸收的可见光红波段(0.6~0.7 μm)和对绿色植物高反射的近红外波段(0.7~1.1 μm)。目前,计算植被指数主要有比值植被指数(RVI)、归一化植被指数(NDVI)和垂直植被指数(PVI)等,其中,归一化植被指数 NDVI 是最常用的一种,其计算式为

$$NDVI = (TM4 - TM3) / (TM4 + TM3) \quad (1)$$

由式(1)得到 NDVI 值范围是(-1,1),在可见光比近红外波段有较高反射的云、水和雪等地物,NDVI < 0;而岩石和裸地在这 2 个波段有近似的反射,NDVI 接近 0;在森林覆盖区 NDVI > 0,一般情况下,NDVI > 0.4 以上时,可认为森林覆盖比较高。试验中,对已经辐射校正和几何校正的 2 期 TM 影像分别进行了主成分分析和 NDVI 计算。

2.3 假彩色合成

通常评价合成影像质量的途径有 2 个,一是采用信息论及数学方法,如信息熵、均方梯度反映光谱信息的偏差、相关系数等客观判断准则;二是通过彩色合成及视觉感官判断的主观判断准则,如目视解译、RGB 彩色合成原理等。本文研究目的是提取森林覆盖信息,波段组合最终效果是要森林信息从背景信息中凸现出来,因而,单波段信息量大小不是决定波段组合方案的主导因素,森林信息和背景信息间的高反差才是本次研究的根本问题。本文采用上述 2 种途径来选择和验证波段组合方案。即通过波段组合相关系数和 RGB 彩色合成原理及视觉感官判断。

以信息量来判断最佳波段组合,以组合的 3 波段标准差之和最大及组合波段间相关系数之和最小为依据。本文将 PCA 得到的第 1 分量 PC 1 作 R(红)波段,NDVI 作为 G(绿)波段,从原始图像中选取某一波段作为 B(蓝)波段进行假彩色合成。对 PC 1、NDVI、TM 1~TM 5 和 TM 7 进行波段间相关性分析。以 2005 年影像为例,计算波段间相关系数,结果如表 1 所示。

表 1 研究区 2000 年影像波段间相关系数矩阵

| 波段 | TM 1 | TM 2 | TM 3 | TM 4 | TM 5 | TM 7 | PC 1 | NDVI |
|------|------------|------------|------------|------------|------------|------------|------------|-----------|
| TM 1 | 1.000 000 | | | | | | | |
| TM 2 | 0.940 965 | 1.000 000 | | | | | | |
| TM 3 | 0.953 307 | 0.944 210 | 1.000 000 | | | | | |
| TM 4 | -0.234 049 | -0.057 976 | -0.289 410 | 1.000 000 | | | | |
| TM 5 | 0.840 829 | 0.847 475 | 0.875 116 | -0.065 956 | 1.000 000 | | | |
| TM 7 | 0.867 437 | 0.832 631 | 0.905 502 | -0.304 700 | 0.938 321 | 1.000 000 | | |
| PC 1 | 0.905 845 | 0.882 372 | 0.934 025 | -0.222 793 | 0.979 089 | 0.978 851 | 1.000 000 | |
| NDVI | -0.722 531 | -0.629 744 | -0.800 839 | 0.761 809 | -0.571 628 | -0.734 942 | -0.701 781 | 1.000 000 |

从表 1 中可以看出, TM 1 ~ TM 3 波段具有很高的相关系数, TM 4 波段比较独立, TM 5 波段和 TM 7 波段相关性较高。因此, 通过相关分析拟选择 TM 2、TM 4 和 TM 5 波段组合。上文提到, 判断波段组合最佳效果信息量不是第一因素, 下面从视觉效果和 RGB 彩色合成来进一步分析最佳波段组合方案。

波段组合主要目的是要使森林信息和其它地物信息产生很大的反差, 使森林信息在影像上凸现出来。计算 *NDVI* 时, 用到 TM 3 和 TM 4 波段; 通过 PCA 得到的第 1 主分量信息中, TM 5 占有较大比重; TM 2 波段位于健康绿色植物的绿色反射率 (0.54 ~ 0.55 μm) 附近, 对森林的绿反射敏感。所以, 从原始影像选择波段时, 可以考虑 TM 1 和 TM 7 (其它组合作为对比)。当影像无云时, 经过组合试验, 采用 PC 1、*NDVI* 和 TM 1 进行 RGB 假彩色合成 (如插页彩片 13 所示), 森林信息和背景被很好地分离开来, 森林表现为绿色, 符合人眼视觉效果; 背景也分成了 3 类, 黑色的水系, 品红色的无林地区域 (如裸地等) 及暗粉红色的少量草/枯草覆盖的土壤 (如枯萎的草丛)。当影像有少量薄云雾 (少于 20%) 时, 云雾会在 TM 1 上表现为噪声, 由于 TM 1 为蓝色波段, 故有少量云雾时, 采用 PC 1、*NDVI* 和 TM 7 进行 RGB 假彩色合成, 有薄云雾覆盖区域表现为墨绿色 (如插页彩片 14 所示)。

2.4 监督分类

本研究需区分 4 种地物: 森林、裸地、草/农地和水域。对选好的样本采用可行性矩阵评价分类模板, 差异百分比 > 85%。从表 2 和表 3 中可以看出, 各个类别的差异百分比都超过了 95%, 这也说明波段重新组合后的 TM 合成影像各类别之间反差较大, 便于样本选取。

表 2 2000 年影像训练样本矩阵

| 类别 | 森林 | 裸地 | 草/农地 | 水域 |
|---------|---------|-------|-------|------|
| 森林 | 187 141 | 0 | 11 | 5 |
| 裸地 | 29 | 6 917 | 12 | 1 |
| 草/农地 | 2 119 | 79 | 5 883 | 4 |
| 水域 | 0 | 0 | 0 | 605 |
| 总数 | 189 289 | 6 996 | 6 041 | 615 |
| 差异百分比/% | 98.9 | 98.9 | 97.4 | 98.4 |

表 3 2005 年影像训练样本矩阵

| 类别 | 森林 | 裸地 | 草/农地 | 水域 |
|---------|---------|-------|--------|-------|
| 森林 | 124 353 | 0 | 35 | 7 |
| 裸地 | 3 | 9 499 | 591 | 0 |
| 草/农地 | 525 | 112 | 34 197 | 1 |
| 水域 | 11 | 0 | 0 | 1 544 |
| 总数 | 124 892 | 9 611 | 34 823 | 1 552 |
| 差异百分比/% | 99.6 | 98.8 | 98.2 | 99.5 |

训练样本选取完后, 对影像进行监督分类, 本文使用最大似然率法, 分类结果如插页彩片 15、16 所示, 其 4 种颜色表示 4 种地物类别。

3 精度评价

由于提取的是森林覆盖信息, 经验表明, TM 5、TM 4、TM 3 波段组合能较好地地区分出森林覆盖区域。本文采用随机抽样的方法, 选取 300 个地面检查点, 每个类别应有一定的数量保证, 这个数量根据每个类别占总数多少来按比例随机分配。首先, 在 TM 影像上对 300 个样点进行目视解译或分类, 得到的人工判读与监督分类结果; 然后进行对比分析; 最后, 得到评价数据, 如表 4 和表 5 所示。

表 4 2000 年影像分类精度评价

| 类型 | 参考点数 | 分类点数 | 正确点数 | 生产者精度/% | 使用者精度/% | Kappa |
|----------------------------------|------|------|------|---------|---------|---------|
| 森林 | 156 | 167 | 150 | 96.15 | 89.82 | 0.787 9 |
| 裸地 | 52 | 41 | 40 | 76.92 | 97.56 | 0.970 5 |
| 草/农地 | 61 | 62 | 44 | 72.13 | 70.97 | 0.635 6 |
| 水域 | 31 | 30 | 30 | 96.77 | 100.00 | 1.000 0 |
| 总体分类精度 = 88.00%; Kappa = 0.810 9 | | | | | | |

表 5 2005 年影像分类精度评价

| 类型 | 参考点数 | 分类点数 | 正确点数 | 生产者精度/% | 使用者精度/% | Kappa |
|----------------------------------|------|------|------|---------|---------|---------|
| 森林 | 142 | 136 | 133 | 93.66 | 97.79 | 0.958 1 |
| 裸地 | 40 | 42 | 37 | 92.50 | 88.10 | 0.862 6 |
| 草/农地 | 89 | 92 | 81 | 91.01 | 88.04 | 0.830 0 |
| 水域 | 29 | 30 | 28 | 96.55 | 93.33 | 0.926 2 |
| 总体分类精度 = 93.00%; Kappa = 0.894 9 | | | | | | |

从表中可看出, 2000 年影像由于森林覆盖高, 裸地和草/农地分类生产者精度不高, 都低于 80%, 但森林生产者精度高, 达到了 96.15%, 总体分类精度为 88.00%, Kappa 指数为 0.810 9; 2005 年影像森林覆盖较低, 各个分类的生产者精度相近, 都达到 90% 以上, 总体分类精度达到 93.00%, Kappa 指数为 0.894 9。从中可以看出, 此方法对森林的分类精度较好, 均超过 92%。

4 分类后处理和森林覆盖图制作

4.1 分类后处理

运用监督分类得到的遥感影像分类结果中, 不可避免地会产生一些面积很小的图斑。无论从专题制图还是从实际应用角度, 都有必要对这些小图斑进行剔除和重新分类。目前, 常用的方法有聚类 (Clump)、过滤 (Sieve)、剔除 (Eliminate) 和邻域分析等。这里采用的是邻域分析法, 用类似于卷积的方

法对图像分类值进行分析,选择 5 像元 × 5 像元的邻域范围和主数值 (Majority) 分析函数。

4.2 森林覆盖图制作

经过分类后处理得到的数据可以用来统计森林覆盖面积、计算一个区域的森林覆盖率、制作森林覆盖专题图及统计森林覆盖变化专题图等。

5 结论

将主成分分析的 PC 1、NDVI 和 TM 某一波段进行假彩色合成,增强了森林信息和背景信息的反差,利用监督分类可以在高精度 (大于 92%) 的情况下将森林信息提取出来,并且对有薄云雾 (小于 20%) 的影像效果也非常明显。从中可以看出本方法较为有效,具有精度高、操作简单且自动解译的特点。

参考文献

- [1] 尹黎明,卢玉东,潘剑君,等. 基于多波段组合技术的土地利用/覆盖信息提取研究[J]. 遥感信息,2005,(5):16-18.
- [2] 李景文. 黑龙江林业[M]. 哈尔滨:东北林业大学出版社,

- 1993.
- [3] 赵英时,等. 遥感应用分析原理与方法[M]. 北京:科学出版社,2003.
- [4] Gong P J, Howarth P. Frequency - based Contextual Classification and Gray - level Vector Reduction for Land - use Identification [J]. Photo. Engin. and Remote Sensing, 1992, 58 (4) : 423 - 437.
- [5] 李石华,王金亮,毕艳,等. 遥感图像分类方法研究综述[J]. 国土资源遥感,2005,(2):1-6.
- [6] Congalton R G. A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data [J]. Remote Sensing Environ. 1991, 37:35-46.
- [7] 刘美玲,齐清文,邹秀萍,等. 基于 RS 对云南边境地区土地覆盖现状及变化研究[J]. 国土资源遥感,2006,(1):75-78.
- [8] 党安荣,王晓栋,陈晓峰,等. ERDAS IMAGING 遥感图像处理方法[M]. 北京:清华大学出版社,2003.
- [9] 吴海平,黄世存. 自动提取土地利用变化信息后处理方法的研究[J]. 国土资源遥感,2005,(1):41-44.

AN APPROACH TO THE EXTRACTION OF REMOTELY - SENSED INFORMATION OF FOREST COVERAGE BASED ON PCA/NDVI

DENG Shu - bin¹, WU Hong - gan², JIANG Tao¹

(1. Shandong University of Science and Technology, Qingdao 266510, China; 2. Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China)

Abstract: With the Daxinganling area as the experimental site, the authors used the PCA (Principal Components Analysis) to get the first weight and the results of the NDVI (Normalized Difference Vegetation Index), and employed a certain band in the LandSat TM1 to 7 to realize the combination of the wave band, which could enhance the contrast between the forest information and the background information. The study also used the method of maximum likelihood to realize the supervised classification of the images, whose accuracy could exceed 92%. This paper presents two experimental investigation images with different quantities of cloud and different extents of forest coverage. This investigation shows that this method can improve the automation and precision in the extraction of the forest coverage information.

Key words: Principal components analysis; Normalized difference vegetation index; Band combinations; Supervised classification; Neighborhood analysis; Maximum likelihood

第一作者简介: 邓书斌(1982 -),男,硕士研究生,主要从事遥感应用研究和 GIS 系统开发。

(责任编辑:肖继春)