

文章编号: 1007-4619 (2003)05-0358-06

# 一种简单加入空间关系的实用图像分类方法

赵红蕊<sup>1</sup>, 阎广建<sup>1</sup>, 邓小炼<sup>1</sup>, 王锦地<sup>1</sup>, 杨 华<sup>1</sup>, 李小文<sup>1,2</sup>

(1. 北京师范大学 遥感与 GIS 研究中心, 资源与环境科学系, 环境遥感与数字城市北京市重点实验室, 北京 100875;

2. Center for Remote Sensing and Department of Geography, Boston University, Boston MA 02215, USA)

**摘 要:** 遥感图像分类是遥感图像处理的一项基本内容,也是遥感应用中关键的一步。为了提高分类的精度,一方面是对光谱信息的合理利用;另一方面,可以加入新的信息源,即进行多源数据处理,并加入地学知识,尤其是对空间信息的利用是至关重要的。但是由于地学知识的复杂性及空间信息利用的难度以及数据源的限制,尚无公认的实用方法。该文提出了一种简单加入空间关系的分类方法,在没有其它数据源的情况下,利用空间关系特性,在分类中构造两个空间关系波段,实现空间约束,部分消除仅依赖光谱数据分类而引起的同物异谱和同谱异物造成的分类错误。简单实用,同时也验证了空间关系在分类中的重要性。

**关键词:** 遥感图像; 分类; 空间信息

**中图分类号:** TP751.1 **文献标识码:** A

## 1 引 言

遥感图像分类是将图像的所有像元按其性质分为若干个类别,为进一步的遥感和应用和理论研究提供信息。目前通常是采用多光谱数据分类。按照是否以先验知识作为确定判别函数和判别准则,遥感图像分类又可分为监督分类和非监督分类。监督分类根据一定数量的已知类别的样本作为训练样本,以其观测值确定判别函数中待定参数,即进行学习或训练,再依据判别准则对该样本的所属类别作出判定。如基于最小错误概率的 Bayes 分类器、子空间分类器等。如果事先没有相关类别的先验知识,而是按照某种相似性准则对像元进行合并或分离,这样的分类方法称为非监督分类,也称聚类(Clustering)。常用的有 K-均值算法(K-means Algorithm),迭代自组织数据分析算法(Iterative Self-Organizing Data Analysis Techniques Algorithm-ISODATA)。这些方法可以见于各商用软件,但都是基于光谱特征的分类方法,图像像元间的空间关系没有受到重视。本文在分析目前各类新的分类方法的基础上,提出了加入空间关系的分类方法,在不增加数据源的情况下,达到多源数据参与分类的效果。

本文在第一部分简单回顾了各种提高分类精度的方法,包括传统的光谱方法及多源数据参与分类的方法;第二部分分析了空间信息在分类中的重要作用;我们提出的加入空间关系的自动分类方法在第三部分介绍,最后利用 MAS 和 ASTER 数据为例验证了新算法。

## 2 提高分类精度的各种方法

为了提高分类精度,很多学者提出了许多方法。这些方法可以归为两类:一是寻求新的方法更为充分地利用光谱信息;二是加入新的数据源,进行多源数据参与的分类研究。

在充分利用光谱信息的前提下,多源数据参与分类可以获得更好的分类结果。邱凯昌、李德仁等<sup>[1]</sup>将光谱信息、GIS 数据和高程数据结合,提出空间对象和像元两种学习粒度,实质是利用了空间信息和多源数据。虽然方法较复杂,但其空间信息的利用较之单纯利用光谱数据的知识挖掘方法,有更好的分类效果。周成虎等将空间关系引入居民地的提取中,在一定程度上避免了纯光谱方法对居民地的多提和漏提<sup>[2]</sup>,提出通过空间和时间的子集分割来进行聚类的方法<sup>[3]</sup>,说明了空间信息的应用前景。

收稿日期: 2002-03-17; 修订日期: 2002-08-03

基金项目: 国家自然科学基金项目 40101020 共同资助、国家重点基础研究发展规划项目(G2000077900)。

作者简介: 赵红蕊(1969—),女,副教授,北京师范大学遥感中心在读博士,主要研究方向:空间信息在遥感中的应用。发表论文十余篇。

遥感图像分类中加入空间信息,是地表结构性的特点的必然要求<sup>[4]</sup>。地学知识的参与也是进一步提高分类精度的关键<sup>[5]</sup>。实践也证明,采用了其它辅助空间数据的分类,其精度都大大高于那些只使用遥感数据的分类<sup>[6,7]</sup>;然而由于对图像中空间结构变化的原因和规律掌握的不足,削弱了空间信息的应用<sup>[4]</sup>。开发能够同时处理遥感图像和其它空间数据的分类算法,是一个非常活跃的研究领域即多源数据集成<sup>[7]</sup>,有望使分类精度提高到新的水平。

### 3 一种简单加入空间关系的实用自动分类方法

针对如何在无法增加数据源的前提下达到多源数据参与分类的效果的问题,本文提出了一种简单加入空间关系的分类方法。其出发点和前述之各种分类方法均不同,它直接利用的是空间自相关这一性质。空间自相关性(spatial autocorrelation),作为地理学第一定律<sup>[8]</sup>,在地学分析中广为应用。它是指空间上越靠近的事物或现象就越相似<sup>[9]</sup>。

按照地学理论,一种地学现象可以用“区域化变量”的空间分布来表征。因为遥感图像是地学现象的影射,所以区域变量必然会在图像上有所反映<sup>[10]</sup>。区域化变量能够客观地描述变量的结构性和随机性,而这种结构性的较常见的含义是指某种地学现象在点  $x$  与  $x+h$  处的状态具有某种程度的自相关<sup>[10]</sup>。

我们提出的方法可以用任何遥感商用软件完成。方法是在分类所用的波段中加入两个人为构造的空间关系波段,以光谱约束方法达到空间约束的目的。公式(1)和(2)是构造空间关系波段图像的公式。

$$B_1(i) = \omega \frac{i-1}{N-1} \quad i = 1, 2, \dots, N \quad (1)$$

$$B_2(j) = \omega \frac{j-1}{N-1} \quad j = 1, 2, 3, \dots, N \quad (2)$$

其中  $B_1(i)$ 、 $B_2(j)$  分别为图像  $B_1$  第  $i$  行、图像  $B_2$  第  $j$  列的灰度值,  $\omega$  为空间约束权重。图像大小为  $N \times N$ 。显然,由公式(1)和(2)分别构造了两幅灰度仅随行变化和仅随列变化的图像。图像的灰度值在 0-1 之间,这是由于在我们研究中参与分类的光谱数据是可见光、近红外大气订正后的反射率,其它形式的数据,需通过归一化处理。

空间约束权重的设置直接影响分类结果,目前我们给出的只是经验值,本例中为  $\omega$  等于 0.7。权重的大小和分类区域的空间结构等多种因素有关,

可通过实验获得最佳的权重。在后续研究中我们将给出定量化权重设置方法。

波段准备好后,可以利用现有的任何商用软件,以各种分类方法分类。

不失一般性,本文以 ISODATA 分类方法为例,分析加入空间关系波段对分类算法的改进。其控制参数有:希望得到的类别数(近似值)、所希望的一个类中的最小数目、类的分散程度的参数、类间距离的描述(如最小距离)、每次允许合并的类的对数、允许迭代的次数。步骤如下:

1) 选取  $N_c$  个类的初始中心  $\{Z_i, i=1, 2, \dots, N_c\}$ 。设有  $k$  个光谱波段(本例为反射率)参与分类,空间关系波段的加入,使维数增加 2,即实际参与分类的波段数为  $k+2$  个。

2) 所有样本  $X$  是  $k+2$  维矢量,  $X$  按如下的方法分到  $N_c$  个类别中的某一类中去:

对于所有的  $i \neq j, i = 1, 2, \dots, N_c$ , 如果  $\|X - Z_j\| < \|X - Z_i\|$ , 则  $X \in S_j$ , 其中  $S_j$  是以  $Z_j$  为中心的类。某个样本  $x$  到第  $j$  类中心的光谱距离为:

$$\sqrt{\sum_{i=1}^k (x_i - z_{ij})^2 + (x_{k+1} - z_{(k+1)j})^2 + (x_{k+2} - z_{(k+2)j})^2}$$

下标表示样本矢量和距离中心矢量的元素号。第  $k+1$  和第  $k+2$  个元素分别对应两个空间维。显然,加入的二维空间以光谱距离的形式实现了空间约束,从而仅使实际空间相邻和相近的样本分为一类。

如果  $S_j$  类中的样本数少于设定最小样本数,则  $N_c = N_c - 1$ , 返回上一步。

3) 以公式(3)计算各类的中心

$$Z_j = \frac{1}{N_j} \sum_{x \in S_j} X \quad (3)$$

$$j = 1, 2, \dots, N_c$$

4) 计算  $S_j$  类内的平均距离

$$D_j = \frac{1}{N_j} \sum_{x \in S_j} \|X - Z_j\| \quad j = 1, 2, \dots, N_c$$

5) 计算所有样本离开其相应的聚类中心的平均距离

$$D = \frac{1}{N} \sum_{j=1}^{N_c} N_j \cdot D_j$$

其中  $N$  为样本总数。

6) 迭代次数检验

7) 计算每类中各分量的标准差  $\delta_{ij}$

$$\delta_{ij} = \sqrt{\frac{1}{N_j} \sum_{x \in S_j} (x_i - z_{ij})^2}$$

式中  $i=1, 2, \dots, k+2$ ,  $k+2$  为样本  $X$  的包括两个空

间维的维数:  $j=1, 2, \dots, N_c$ ,  $N_c$  为类别数;  $k=1, 2, \dots, N_j$ ,  $N_j$  为  $S_j$  类中的样本数;  $x_{ik}$  为第  $k$  个样本的第  $i$  个分量;  $z_{ij}$  为第  $j$  个聚类中心  $Z_j$  的第  $i$  个分量;

8) 对于每一个聚类  $S_j$ , 找出标准差最大的分量  $\delta_{j\max}$

$$\delta_{j\max} = \max(\delta_{1j}, \delta_{2j}, \dots, \delta_{kj}, \delta_{xj}, \delta_{yj})$$

$$j = 1, 2, \dots, N_c$$

在  $x$  和  $y$  两个空间维上, 分量标准差  $\delta_x$  和  $\delta_y$  表示空间维分量, 它们随空间距离的增加而增加, 从而实现了分类时的空间约束。

如果下面两个条件有一个成立, 则把  $S_j$  分裂成两个聚类。

条件 1:  $\delta_{j\max}$  大于所设标准差阈值, 且  $D_j > D$  且  $D_j > 2 \cdot (\theta_N + 1)$ 。  $\theta_N$  为所设一类样本最小数目阈值。

条件 2:  $\delta_{j\max}$  大于所设标准差阈值, 且  $N_c < \frac{k}{2}$ ,  $k$  为所设类别数阈值。

9) 计算所有聚类中心间的两两距离

$$D_{ij} = \|Z_i - Z_j\|, i = 1, 2, \dots, N_c - 1,$$

$$j = i + 1, \dots, N_c$$

并对小于所设阈值的类按既定规则进行合并。

10) 若迭代次数大于阈值或迭代中的参数的变化在阈值以内, 则迭代结束, 否则转向第 2 步进行迭代。

可见, 构造的两个空间关系波段以灰度随距离的变化实现了分类时的空间约束, 通过参与类内、类间距离、标准差的计算, 将空间约束融入分类过程。从而保证同类地物空间上的相邻或相近, 满足空间自相关特性。

## 4 利用 MAS 数据的分类实例

作为简单实例, 本文所用的数据是 MAS (MODIS AIRBORNE SIMULATOR) 图像, 成像时间是 1999-03-19。地点是美国 Wisconsin 的 Madison。该图像共有 50 个波段, 可见光、近红外共有 9 个波段, 我们选取了第 2、3、7 波段参与分类(我们亦用其它波段组合进行分类, 所得结果相似)。其峰值波长分别是: 0.5540、0.6580、0.8740。空间分辨率是 50m。为了实验, 我们选择了 Madison 机场所在的  $40 \times 40$  的一块小区域, 因为机场跑道清晰可辨, 可以帮助我们在对地面情况缺乏了解的情况下判断分类的结果。另外, 为了说明问题, 我们分类很细, 共分了 60 类, 以便于比较分类结果是否反映了真实的地面空间分布模式。

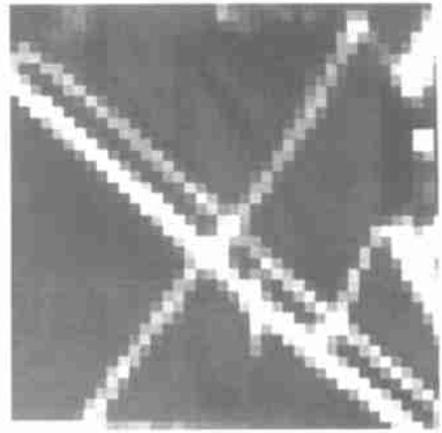


图 1 分类区域的彩色合成图

Fig. 1 Color-combined image of the classified area

图 1 是该区域的彩色合成图, 所用波段即为波段 2、3、7。采用的分类方法是 ISODATA 方法, 在分类中加入了两个空间关系波段, 也即构造了两幅和可见光、近红外同样大小 ( $N \times N$ , 这里  $N=40$ ) 的图像  $B_1$  和  $B_2$  参与分类,  $B_1$  和  $B_2$  各像素的值分别用公式(1)和公式(2)给定。以此来近似地使两点间空间距离的变化等于其灰度值的变化, 达到分类时空间约束的目的。通过空间约束, 分类结果既保持了光谱特征的一致性, 又在空间上相邻或相近, 对局部分类精度的提高是显而易见的(图 2、3); 但是对于空间迥异而同类的地物, 加入空间关系后会作为不同的类别分出来。这就需要进一步处理(见本文后续部分)。如下面将要讨论的分类中涉及到的飞机跑道, 加入空间关系后使跑道分成了若干段, 要得到整个的跑道, 需进行类的合并。我们发现, 这种考虑局部空间自相关的分类结果和景观研究中的最小功能模块类似。每一类别均具有一定的空间上的“同质性”。而每一个最小功能模块都有一个相应的作用区域——最小空间部分。尤其是在景观研究中需要寻找这样的一定空间上具有“同质性”的区域。在分类中加入空间信息是值得探讨的研究方向。

## 5 结果分析与讨论

加入空间关系的分类结果和没有加入空间关系, 仅依传统的光谱波段的分类结果比较, 加入了空间关系后清晰地反映了图像的空间分布模式(见图 2、3)。因为所选区域有机场跑道, 可以帮助判断分类的结果。加入空间关系后, 跑道非常清晰地划分出来了。当然, 由于空间关系的限制, 跑道被分成了

若干段。任意选定跑道上的一类,比较两种分类方法的结果。图 4 是加入空间关系的分类结果中的一类,图 5 是未加空间约束的 ISODATA 分类方法的分类结果中的一类。机场跑道和周围地物的反射率差别较大,仅靠光谱值还是可以区别开的,只是区别的精度问题。如果光谱值差别较小,而空间差异较

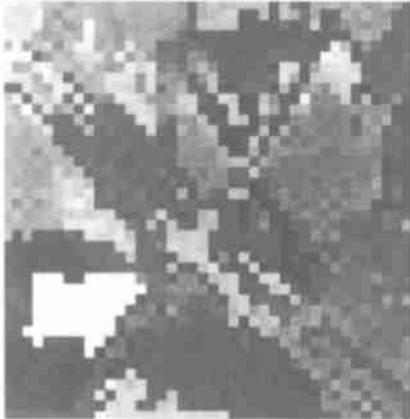


图 2 加入空间关系的分类结果

Fig.2 Classification result with spatial relation participated



图 3 传统分类方法的分类结果

Fig.3 Classification result with traditional method

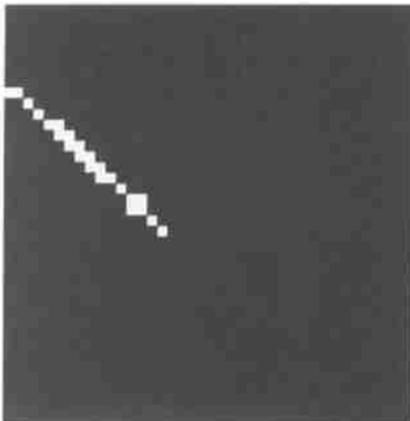


图 4 加入空间关系分类中一类

Fig.4 One of the classes derived from new method

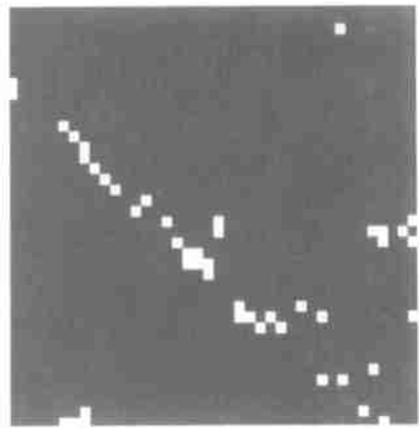


图 5 传统分类方法中一类

Fig.5 One of the classes derived from traditional method

大的地物,加入空间关系的分类方法就更显示其优势。最重要的,是该方法简单、易行,可以利用现有的任何软件。如果将该方法和知识挖掘、信息熵判断等方法结合起来,动态判定空间关系的权重,会收到更好的分类效果。

加入空间关系分类的一个明显不足是把同类地物(如跑道)分成了不同的类。所以,这种方法适宜于寻找一定空间上具有“同质性”的区域,表达地物的空间分布模式,用于传统意义上的分类仅需进一步根据光谱特性进行类的合并。首先以参与分类的光谱波段计算各类的平均反射率。再给定合适空间光谱差阈值,进行类的合并。下面是一个简单的处理步骤:

(1) 对于每个类的  $k$  个光谱波段取类均值,并将均值赋给参与均值计算的每个像元。得到  $k$  个类反射率均值图。



图 6 可见光、近红外第一主分量

Fig.6 The first component of VNIR bands



图 7 本文加入空间关系的分类结果

Fig.7 Classification result with spatial relation participated



图 8 传统 ISODATA 的分类结果

Fig.8 Classification result with traditional ISODATA method



图 9 图 7 和图 8 局部放大图,左、右

图分别对应图 7、图 8 标识区域

Fig.9 Local parts of Fig.7 and Fig.8. The left and right

parts correspond with Fig.7 and Fig.8 respectively

(2) 以  $k$  个均值图用传统光谱方法分类。经过 1 和 2 两步得到的分类结果和传统分类方法相比,由于均值图包含了空间自相关信息,大大减少了“椒盐现象”。我们截取了北京首都机场附近  $400 \times 400$  的 aster 图像。图 6 是其可见光、近红外第一主分量、图 7、图 8 分别表示本文加入空间关系和传统 ISODATA 的分类结果图。明显可以看出加入空间关系的分类方法“椒盐现象”要少得多。图 9 是图像右下角跑道部分放大图(见图 7 和图 8 椭圆标识部分),左图为本文所述方法,右图对应图 8,为 ISODATA 的分类结果。加入空间关系分类的优势一目了然。重要的是,本文所述方法没有对空间数据源的依赖,且简单易用。可以说是以简单的方法证明在分类中加入空间关系是必要的,也是可行的。它可以减少由于同物异谱和同谱异物造成的分类错误。在其它领域(如景观研究)的应用也值得在此基础上继续探讨。

## 参 考 文 献 (References)

- [1] Di K C, Li D R, Li D Y. Study Of Remote Sensing Image Classification Based On Spatial Data Mining Techniques [J]. *Journal of Wuhan Technical University of Surveying and Mapping*. 2000, 25, (1): 42-48. [邱凯昌、李德仁、李德毅. 基于空间数据挖掘的遥感图像分类方法研究. 武汉测绘科技大学学报, 2000, 25, (1): 42-48.]
- [2] Zhou C H, et al. The Geographic Comprehension And Analysis of Remote Sensing Image [M]. Beijing: Science Press, 2000. [周成虎等. 遥感影像地学理解与分析[M]. 北京: 科学出版社, 2000.]
- [3] Zhou C H, Zhang J T. Entropy-Based Model For Geo-Data Mining [5]. *Journal of Image and Graphic*. , 1999, 4(11): 946-951. [周成虎, 张健挺. 基于信息熵的地学空间数据挖掘模型. 中国图像图形学报, 1999, 4(11): 946-951]
- [4] Dong Q, Feng L X. The Use of The Variogram in Remotely Sensed Images [J]. *Remote Sensing Technology and Application*. 1997, 12, (1): 8-13. [董庆、冯林新, 变差函数在遥感图像处理中的应用. 遥感技术与应用, 1997, 12(1): 8-13.]
- [5] Luo J C, Zhou C H, Yang Y. Radial Basis Function Map Theory Based Remote Sensing Image Classification Modal [J]. *Journal of Image and Graphic*. , 2000, 5(A)(2): 94-99. [骆剑承, 周成虎, 杨艳. 基于径向基函数(RBF)映射理论的遥感影像分类模型研究[J]. 中国图像图形学报, 2000, 5(A)(2): 94-99.]
- [6] Zhu S L, Zhang Z K. The Aquirement And Analysis of Remote Sensing Image [M]. Beijing: Science Press, 2000. [朱述龙、张占陔. 遥感图像获取与分析[M]. 北京: 科学出版社, 2000.]
- [7] Tian Q. An Algorithm For Spatial Data Integrated Classification Based on Goodall'S Affinity Index [J]. *Journal of Remote Sensing*. 1999, 3(3): 187-192. [田青. 基于GOODALL 相近指数的遥感图像和其它空间数据综合分类方法[J]. 遥感学报, 1999, 3(3): 187-192.]

- [8] Tobler W R. A Computer Movie Simulating Urban Growth In The Detroit Region [J]. *Economic Geography*, 1970, **46**: 234—240. [10] Guan Z Q. Study on the Theory of Remotely Sensed Imagery Analysis Based on Quotient Space Description [D]. Dissertation, 1995. [关泽群. 商空间下的遥感图像分析理论探讨[D], 学位论文, 1995.]
- [9] Wu J G. Landscape Ecology [M]. Beijing: Higher Educational Press, 2000, 119—131. [邬建国, 景观生态学[M], 北京: 高等教育出版社, 2000. 119—131.]

## A Classification Method Based on Spatial Information

ZHAO Hong-ru<sup>1</sup>, YAN Guang-jian<sup>1</sup>, DENG Xiao-lian<sup>1</sup>, WANG Jin-di<sup>1</sup>, YANG Hua, LI Xiao-wen<sup>1,2</sup>

(1. *Research Center for Remote Sensing, Dept. Geography and Beijing Key Laboratory for Remote Sensing of Environment and Digital Cities, Beijing Normal University Beijing 100875, China;*

2. *Center for Remote Sensing, Boston University, Boston MA 02215, USA*)

**Abstract:** Remote sensing image classification is one of the basic components in remote sensing image processing, which is the key stage in remote sensing applications. There are two ways to get a higher precision: one is to make full use of spectral information, and another is to add other new information. The later is called multi-information processing. Spatial information is one of the most important information in these kinds of multi-information. Since lack of information is always the key problem in remote sensing image processing, we propose a spatial information based method to cope with this problem. What we used is not the spatial data but the spatial characteristics.

The main advantage of our method is that it can work without any other data sources. We take advantage of spatial autocorrelation law, which is the first law of geography. The law is widely used in geo-analyses. It denotes that the closer in spatial, the similar the objects or phenomenon is. Two spatial bands are constructed based on spatial autocorrelation law and a spatial classification result is produced. These two additional bands containing spatial location information are used in classification. One real class may be divided into many classes that are not similar in space. But every class has similar spectral and spatial characteristics. Based on the classification results, further processing should be done. First, we replace every pixel value of every spectral band with the average spectral value of its class. Then a traditional classification method, such as ISODATA method, can be used. With these steps, we get the last classification results.

The MAS and ASTER images are used to prove our method. The classification results show that some miss classified ground covers in the general spectrum based classification method are thrown off by our method. In other words, our method is resistant to the problem of having apparent same spectrum but different ground cover. Another distinct advantage of our method is about the removal of salt and pepper noise from images. It may be also helpful in other research area, such as landscape research. Our method also proves the importance of spatial information in remote sensing image classification. It gives an easy way to add spatial constraint to classification process. The new classification method reveals its simpleness and practicality.

**Key words:** remote sensing image; classification; spatial information