

文章编号:1001-9014(2004)04-0299-04

# 一种基于空间连续性的高光谱图像分类方法

耿修瑞, 张霞, 陈正超, 张兵, 郑兰芬, 童庆禧  
(中国科学院遥感应用研究所 遥感信息开放实验室, 北京 100101)

**摘要:**利用地物在空间上分布连续性这一信息,提出了一种高光谱图像先分块、再分类的分类方法,改进了通常意义下的分类只考虑光谱信息这一缺陷,通过对2000年我们在日本长野南牧村,采用上海技术物理研究所研制的推扫式成像光谱仪(PHI)获得了80个波段的高光谱数据进行实验,验证了我们算法的优越性。

**关键词:**最小距离法;空间连续;高光谱;聚类  
**中图分类号:**P237 **文献标识码:**A

## CLASSIFICATION ALGORITHM BASED ON SPATIAL CONTINUITY FOR HYPERSPECTRAL IMAGE

GENG Xiu-Rui, ZHANG Xia, CHEN Zheng-Chao, ZHANG Bing,  
ZHENG Lan-Fen, TONG Qing-Xi

(Laboratory of Remote Sensing Information Science, Institute of Remote Sensing Applications,  
Chinese Academy of Sciences, Beijing 100101, China)

**Abstract:**Based on the spatial continuity of ground objects, a new classification algorithm was proposed by two steps, namely blocking images followed by further classification, which improved the disadvantage of the general classification algorithms constrained by only spectral information. The new algorithm proposed in this study was validated by the 80-band hyperspectral image acquired by the push broom hyperspectral imager (PHI) in Minamimaki of Japan, 2000. The PHI sensor was developed by the Shanghai Institute of Technical Physics. It turned out that our algorithm embodied the advantage in classifying hyperspectral image.

**Key words:** minimum distance; spatial continuity; hyperspectra; clustering

### 引言

传统的高光谱图像分类方法,无论是最近邻法、最大似然法、SAM方法,还是混合决策树法<sup>[1,2]</sup>神经网络方法<sup>[3]</sup>是先选取训练区进行训练,由每个训练区得到一个平均光谱,然后将每一类得到的平均光谱跟需要分类的高光谱图像一一比较其相似性,最终得到分类结果。这样的处理方法难免会造成‘麻点’现象,即同一地块中夹杂着不该有的其它类别,从而造成分类效果下降,该现象一般是由噪声和地物分布在数据获取尺度上的不均一性所引起的。由于所得到的高光谱图像一般都是含有噪声的,该噪声一般情况下可近似为白噪声。又因为原始图像中每条光谱曲线可以认为是它的具有统计意义平均

光谱的一个样本,所以只要训练样本足够大,可以认为每个训练区得到的平均光谱更加逼近真实光谱。这样,拿这些比较精确的光谱跟含有噪声的不具有统计意义的单个原始光谱比较相似性进行分类时,就会产生‘麻点’现象,继而引起分类精度的降低。

本文提出了一种利用光谱信息兼顾空间信息的高光谱图像分类方法,研究表明该方法能够保持地块的连续性和均一性,同时有力地克服了‘麻点’现象,从而大大提高分类的效果。

### 1 利用空间信息的高光谱图像分类方法

传统的高光谱图像分类方法大多注重地物在光谱维上的差异,而忽视了它们空间维上的结构。在利用空间信息方法,赵永超提出的SRSSH算法<sup>[4]</sup>已

收稿日期:2003-02-26,修回日期:2003-10-10

基金项目:中科院遥感所知识创新课题资助项目(CX020018)

作者简介:耿修瑞(1977-),男,河南封丘人,博士研究生,主要从事高光谱图像处理。

Received date: 2003-02-26, revised date: 2003-10-10

经注意到了空间信息的重要性,张兵提出了一种时空辅助下的植被分类构想<sup>[5]</sup>,但需要借助于高地面分辨率的航片来完成分类.一般情况下,地物在空间上都是连续分布的,或者说如果某点处为 A 类地物,那么在此点的领域上是 A 类地物的可能性最大,这是一个非常重要的信息.利用好这个信息是本文去除‘麻点’现象、提高分类精度的关键所在.本文提出了一种先空间领域自动聚类、然后再分类的算法,得到了很好的分类效果.

用  $D$  代表需要分类的高光谱数据立方体,则  $D_{ij}$  表示在空间坐标  $(i, j)$  处的光谱  $(i = 1, 2, \dots, M; j = 1, 2, \dots, N)$ ,其中  $M, N$  分别表示需要分类的高光谱图像在空间维上的行、列数.

我们引入一个  $M$  行、 $N$  列的分块矩阵  $S$  以及一个分块阈值  $\mu$ ,其中的元素  $S_{ij}$  表示在空间坐标  $(i, j)$  处  $D_{ij}$  所属的块数.用  $F(x, y)$  表示两条光谱曲线光谱  $x$  和  $y$  的相似程度(比如某种距离),当  $F(x, y) \leq \mu$  时,  $x$  和  $y$  的相似程度比较大,则把  $x$  和  $y$  归为同一块;反之,当  $F(x, y) > \mu$  时,  $x$  和  $y$  的相似程度比较小,则  $x$  和  $y$  属于不同的块.具体的处理步骤如下:

1) 令  $S_{11} = 1$ ,即原始高光谱图像第一行,第一列的像元属于第一类.

2) 如果  $F(D_{1j}, D_{1,j-1}) \leq \mu$ , 则  $S_{1,j-1} = S_{1j}$ , 否则,  $S_{1,j-1} = S_{1j} + 1 (j = 2, 3, \dots, N)$ .

3) 当  $j = 1$  时, 令  $t = \min(F(D_{i1}, D_{i-1,1}), (D_{i1}, D_{i-1,2}), F(D_{i1}, D_{im})) = t$ , 而  $t \leq \mu$  时, 则  $S_{i1} = S_{im}$ ; 否则  $S_{i1} = \max(S) + 1$ . 当  $j \neq 1$  时, 令  $t = \min[F(D_{ij}, D_{i-1,j-1}), F(D_{ij}, D_{i-1,j}), F(D_{ij}, D_{i-1,j+1}), F(D_{ij}, D_{i,j-1})], F(D_{ij}, D_{im}) = t$ , 而  $t \leq \mu$  时, 则  $S_{ij} = S_{im}$ ; 否则,  $S_{ij} = \max(S) + 1 (j = 2, 3, \dots, N)$ . ( $i = 2, 3, \dots, N$ )

4) 求与矩阵  $S$  中数值相同的块对应的数据立方体  $D$  的所有相应块的光谱平均值.例如,如果在矩阵  $S$  中有且仅有  $S_{ij} = S_{i,j+1} = S_{i+1,j} = k$ . 那么  $D$  中相应位置的信应改变为  $D_{ij} = D_{i,j+1} = D_{i+1,j} = \frac{D_{ij} + D_{i,j+1} + D_{i+1,j}}{3}$ .

5) 对所得到的  $D$  再用适当的方法进行分类.

用图 1 来说明上面的分块步骤,图中每个方格表示一个像元,代表此处的光谱.自左到右、自上至下遍历整个空间区域.当遍历到红色的方格(像元)时,它需要和它的领域里已经遍历过的方格(即 4 个蓝色的方格)相比较,根据他们之间的相似性决定红色方格应和哪一个蓝色方格属于同一块或者跟

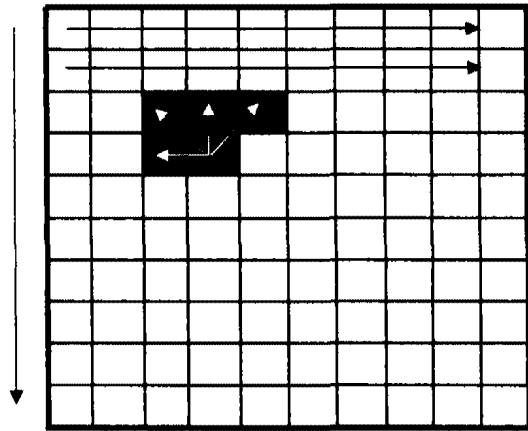


图 1 分块步骤示意图

Fig. 1 The sketch map of blocking process

他们都不同属同一块.遍历整个空间范围之后,给所有属于同一块的光谱取平均,这样就得到原始高光谱图像的分块图像  $E$ ,在  $E$  中每个块中的像元的光谱都是相同的(即本块的平均光谱),再对  $E$  按适当的分类方法进行分类,将得到没有‘麻点’效应的分类结果.

由上面的处理可看出,  $E$  中的每个像元的光谱不再是带有噪声的原始光谱,而是每个块中所有光谱的平均,由于我们块的划分是严格的,既要求空间相邻,又要求光谱相似,因此可以认为属于同一块的像元必定属于同一类,对多数高光谱图像,这样的假设是合理的.经过分块处理之后的光谱在很大程度上压制了噪声,从而得到更为准确的具有统计意义的光谱.再对  $E$  进行分类的时候,将不是拿训练区的平均光谱跟原始的带有噪声的光谱比较,而是跟每个块的平均光谱相比较,这样做的好处是不言而喻的.

## 2 实验与分析

2000 年我们在日本长野南牧村某农作物区利用上海技术物理研究所研制的推扫式成像光谱仪 (PHI) 获取了 80 个波段的高光谱数据,先利用平场域法行了反射率转换,然后截取其中  $200 \times 200$  的图像作为实验区域.因为非植被地物一般来说是容易分类的,本文仅对该区域内的植被进行处理,因此首先采用 NDVI 阈值法将植被部分分离出来,再进行下面的处理.

本文以光谱间的欧氏距离作为它们之间相似性的量度,即把前面分块步骤中所提到的表征光谱间相似程度的  $F(x, y)$  用  $d_{xy}$  代替:

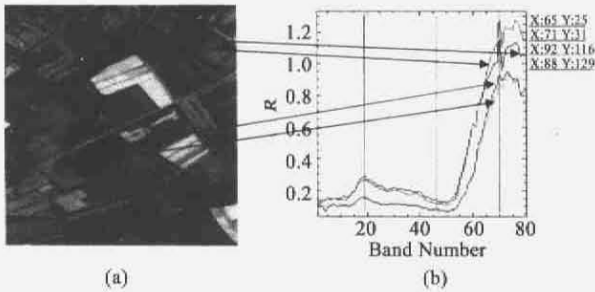


图2 原始图像及光谱(a)为该区域的假彩色合成图(波段70,46,19), (b)为从图1(a)相应位置上提取的反射率光谱

Fig. 2 The original image and the spectra extracted from it (a) false color composite image (band 70, 46, 19) (b) reflectance spectra extracted from corresponding position in fig. 2 (a)

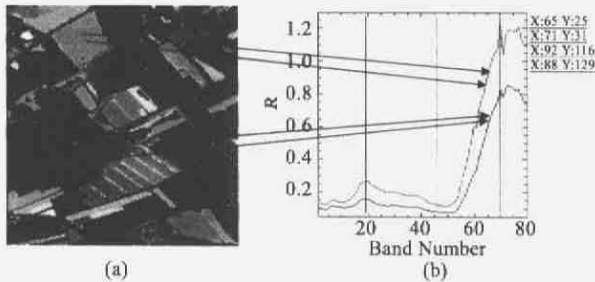


图3 分块图像及光谱(a)为该区域植被经过我们算法处理之后的假彩色分块图(波段70,46,19), 其中 $\sigma_1 = 0.12$  (b)为图像相应位置上的反射率光谱

Fig. 3 Blocking image and its spectra (a) false color composite blocking image (band 70, 46, 19) after processing, where  $\sigma_1 = 0.12$  (b) reflectance spectra extracted from corresponding position in fig. 3 (a)

$$d_{xy} = \sqrt{(b_{x1} - b_{y1})^2 + (b_{x2} - b_{y2})^2 + \dots + (b_{xn} - b_{yn})^2} \quad (1)$$

其中,  $n$  是波段数(这里  $n = 80$ ).  $x = (b_{x1}, b_{x2}, \dots, b_{xn})$  和  $y = (b_{y1}, b_{y2}, \dots, b_{yn})$  为光谱矢量.

这里要用到2个阈值 $\sigma_1$ 和 $\sigma_2$ (其中 $\sigma_1$ 相当于前面的分块阈值 $\mu$ ), 在遍历图像时, 规定当 $d_{xy} < \sigma_1$ 时,  $x$ 和 $y$ 属于同一块; 否则, 他们不属于同一块. 在给图像分类时, 规定当 $d_{xy} < \sigma_2$ 时,  $x$ 和 $y$ 属于同一类, 否则它们不属于同一类. 其中 $\sigma_1$ 和 $\sigma_2$ 的选取依赖于具体的高光谱图像, 当所选区域不同地块的植被有明显的界限时(比如田埂、道路),  $\sigma_1$ 一般可以选的比较小; 否则,  $\sigma_1$ 应比较大. 在使相邻的距离最小的两条光谱分为不同块的同时,  $\sigma_1$ 越大, 得到的分类结果越好. 一般情况下,  $\sigma_1$ 和 $\sigma_2$ 的选取应满足 $\sigma_1 > \sigma_2$ .

比较图2和图3中的光谱曲线, 可以看出, 经过

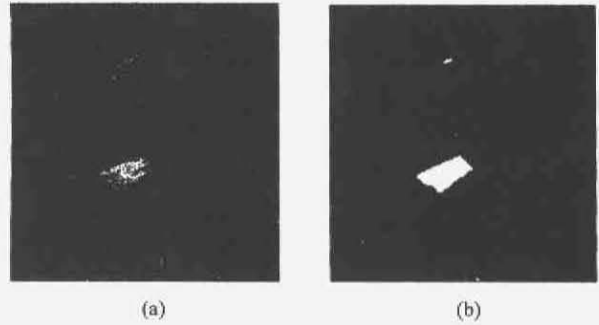


图4 提取地块结果(a)为用最小距离法直接从原始高光谱图像所提取的某种植物(架豆)的结果 (b)为经过我们算法处理后的分块图像所提取的结果, 其中 $\sigma_2 = 0.06$

Fig. 4 The result of extracting certain species (a) the result of directly extracting certain vegetation (soybean) by minimum distance algorithm from the original hyperspectral image (b) the corresponding result by our algorithm, where  $\sigma_2 = 0.06$

算法处理后的图像, 噪声和地物分布的不均匀性被明显地改善了, 得到了较为光滑的光谱和经过先验知识验证的均一的地块(比如, 图3中部棕色地块均为架豆), 从而也更接近地物的真实光谱. 图4显示出算法在提取均一地块的优越性. 图5则进一步验证了我们的观点. 从图5中可以看出直接对原图像进行分类误分、麻点现象较为明显(比如, 均一的架豆地块被分成了好几类), 但经过我们分块算法处理后再对图像进行分类, 很好地避免了这一现象.

本文算法对于所研究区域不同地物在交界处且在光谱上有明显的差别时效果尤其明显, 但是在相邻地块不同地物的光谱非常接近或者差别不大的情况下, 有可能会把它们归为一块, 从而最终划分为一类, 也就是说, 当不同地物在空间上也体现一定的连

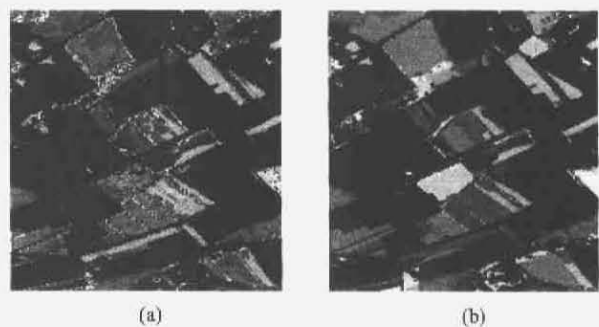


图5 无监督分类结果(a)为利用ENVI上的IsoData方法对原始高光谱图像进行无监督分类的结果 (b)为对我们算法处理过之后的图像做相应处理的结果

Fig. 5 The unsupervised classification result (a) the unsupervised classification result by using IsoData algorithm in original hyperspectral image (b) the corresponding result by our algorithm

续性的时候,利用我们的算法就会导致误分现象;同时,由于我们的算法更侧重于地物的空间连续性信息,即使是同一地块的同一种地物,如果不能保证这种连续性,即它们光谱差别比较大的时候,采用我们的算法就可能会把它们归为不同的类别,比如,同一块地不同长势的植被就属于这种现象.

### 3 结语

对于任何问题,信息量的增加一般会有助于问题的解决.在利用光谱信息进行分类的同时,这里又引入了一个便于操作的空间信息,即空间连续性.实验证明,在对高光谱图象进行分类的时候,该方法不失为一种行之有效的方法.

当然,由于最小距离法以及 IsoData 法其本身没有充分利用地物的光谱信息,特别是各个波段之间的相对关系,由此而引起的误分现象是在所难免的,但是这难以掩盖运用空间连续性这一信息对分类带来的好处.

### REFERENCES

- [1] XIONG Zhen, TONG Qing-Xi, ZHENG Lan-Fen. High-rank artificial neural network algorithm for classification of hyperspectral image data[J]. *Journal of Image and Graphics*, (熊桢,童庆禧,郑兰芬.一种用于高光谱遥感图像分类的高阶神经网络算法. *中国图象图形学报*), 2000, 5(3): 196—201.
- [2] Carlotto Mark J. Spectral shape classification of landsat thematic mapper imagery[J]. *Photogrammetric Engineering & Remote Sensing*, 1998, 64(9): 905—913.
- [3] XIONG Zhen, ZHENG Lan-Fen. Hierarchical neural network classification algorithm[J]. *Journal of Survey and Mapping* (熊桢,郑兰芬.分层神经网络分类算法. *测绘学报*), 2000, 29(3): 229—234.
- [4] ZHAO Yong-Chao, TONG Qing-Xi, ZHENG Lan-Fen, et al. A Kernel Adaptive Filter (SRSSH) and Quality Improvement Method for Hyperspectral Image on the Base of Spectral Dimension Recognition and Spatial Dimension Smoothing According to CSAM[C]. *SPIE: SPIE 2nd International Symposium on Multispectral Image Processing and Pattern Recognition*, 2001, 4552, 230—236.
- [5] ZHANG Bing, ZHANG Xia, LIU Liang-Yun, et al. Spectral unmixing and image classification supported by spatial knowledge[C]. *Proceedings of SPIE*, 2003: 4897: 279—283.