

基于 LSMM 和改进的 FCM 提取城市植被覆盖度 ——以北京市海淀区为例

胡姝婧, 胡德勇*, 赵文吉

(三维信息获取与应用教育部重点实验室, 北京 100048; 资源环境与地理信息系统北京市重点实验室, 北京 100048;
首都师范大学资源环境与旅游学院, 北京 100048)

摘要: 植被是城市生态系统的重要组成部分, 及时获取植被覆盖信息对城市生态环境监测具有重要意义。利用中分辨率 Landsat TM 遥感数据, 采用线性光谱分解模型 (LSMM) 开展城市植被覆盖度提取; 同时, 通过改进训练样本选择方式, 在最小噪声变换 (MNF)、像元纯净指数分析 (PPI)、N 维可视化分析基础上得到端元样本, 再运用模糊 C-均值 (FCM) 获取植被覆盖度; 最后以高分辨率 SPOT5 遥感数据对两种方式的提取结果进行精度检验。结果显示, 基于 LSMM 和改进的 FCM 提取的城市植被覆盖度与检验数据相关系数分别为 0.8252 和 0.9381, 后者可以较好地处理其他要素的非线性影响, 因而具有较高精度。

关键词: 植被覆盖度; 混合像元分解; 线性光谱混合模型; 改进的 FCM

Extract urban vegetation coverage based on LSMM and improved FCM: a case study in Haidian District

HU Shujing, HU Deyong*, ZHAO Wenji

Key Laboratory of 3D Information Acquisition and Application of Ministry of Education, Beijing 100048, China

Key Laboratory of Resources Environment and GIS of Beijing Municipal, Beijing 100048, China

College of Resource Environment and Tourism, Capital Normal University, Beijing 100048, China

Abstract: Vegetation is an important part of urban ecosystem; therefore timely access to vegetation coverage information is of great significance for monitoring urban ecological environment. Linear spectral mixture model (LSMM) was carried out for urban vegetation coverage extraction using medium-resolution Landsat TM remote sensing data. Meanwhile, the fuzzy c-means (FCM) method was chosen to extract vegetation coverage by improving the training sample selection method to obtain the end-member sample based on minimum noise transform (MNF), pixel purity index analysis (PPI), and N-dimensional visualization analysis. Finally, high-resolution SPOT5 remote sensing data extracted in two ways were used to carry out the accuracy test for vegetation coverage. The results showed that the correlation coefficients between the inspection data and LSMM-based and improved FCM-based data were 0.8252 and 0.9381, respectively. It indicated that the improved FCM-based method with higher accuracy can better eliminate the nonlinear effect of other elements.

Key Words: vegetation coverage; pixel unmixing; linear spectral mixture model; improved fuzzy c-means

植被是城市生态系统的子系统, 植被信息的准确提取是区域生态环境质量评价的重要基础。利用遥感技术可以及时、动态地提取城市植被覆盖信息。

估算植被覆盖度的方法大体为两种: 植被指数法和混合像元法, 前种方法有赖于特定区域的实测数据和卫星传感器, 推广应用受诸多限制^[1]; 而后种方法可以提高遥感图像的分类精度^[2], 得到较为准确的覆盖度

基金项目: 遥感科学国家重点实验室开放基金资助项目; 国家科技支撑计划资助项目 (2007BAH15B02); 国家 863 计划资助项目 (2009AA12Z-14611)

收稿日期: 2009-08-10; 修订日期: 2009-11-11

* 通讯作者 Corresponding author. E-mail: deyonghu@163.com

信息。目前,用于混合像元法的模型共 5 类:线性模型、概率模型、几何光学模型、随机几何模型以及模糊模型^[3],其中线性模型和模糊模型使用较为广泛。Small 先后提出在主成分变换及最小噪声分离变换基础上,采用线性光谱模型计算城市植被覆盖度^[4-5]。但线性模型本身存在不足,它认为某一像元的光谱反射率为各组分光谱反射率的简单相加,而多数情况下为非线性组合。另外,目前常采用带全部约束或部分约束条件的最小二乘法^[6]进行线性光谱分解,但该方法对局部噪声、大气效应、环境辐射的影响十分敏感而导致分解结果误差较大,且带部分约束条件的最小二乘法易出现覆盖度小于“0”或大于“1”的不合理值^[7]。模糊 C-均值(FCM)是一个聚类算法,可用于遥感监督分类,霍东民指出当图像上所有类别都被定义并且参与训练时,FCM 可得到混合像元类别组成的精确估计^[8]。但监督分类中人工选择训练样本要求对研究区有充足先验知识,并存在一定主观性。

本文针对研究现状,利用 2007 年 5 月的 T M 数据,一方面采用线性光谱分解模型(LSMM)得到北京市海淀区植被覆盖度;另一方面,对训练样本选择方式进行修改,在最小噪声变换(MNF)、像元纯净指数分析(PPI)、N 维可视化分析基础上得到端元,由端元所在栅格生成监督分类的训练区,再运用 FCM 算法得到植被覆盖度;最后辅以同期 SPOT5 数据对两种方式的提取结果进行精度检验。

1 研究区和数据

海淀区位于北京市区西北部,处于 39°53′—40°09′N,116°03′—116°23′E,面积 426km²。地势西高东低,西部为海拔 100m 以上的山地;东部和南部为海拔 50m 左右的平原。它是典型的城乡结合带,中心城区建筑密度较高,西北部山区植被茂密,多种土地利用类型并存。海淀区地理位置如图 1 所示。

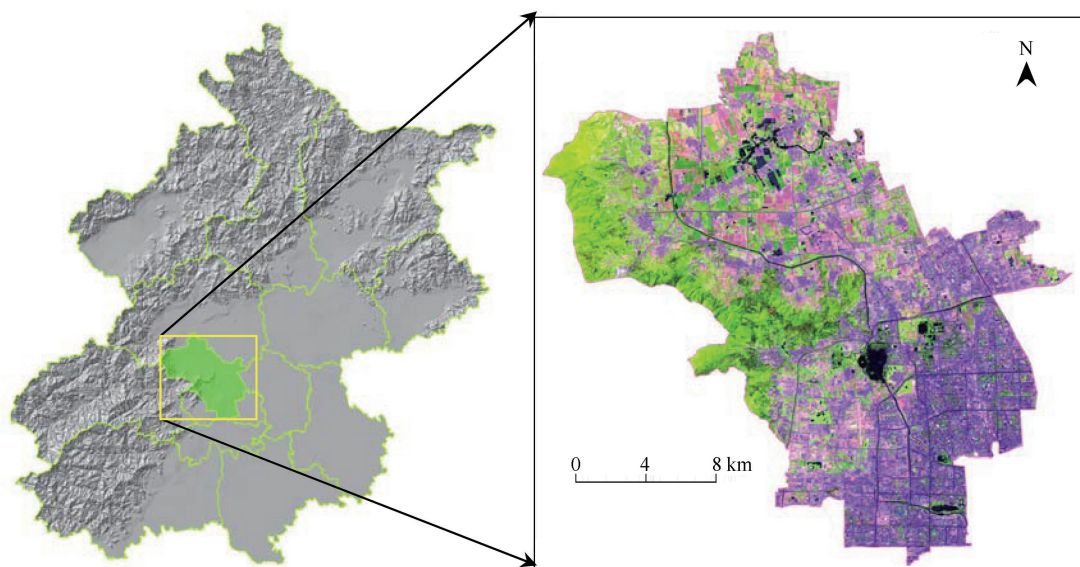


图 1 海淀区地理位置示意图

Fig.1 Geographical location of Haidian District

本文采用中分辨率影像 Landsat T M 提取植被覆盖度,影像获取时间为 2007 年 5 月,分辨率为 30m。另有 2007 年 5 月的 SPOT5 影像,分辨率为 2.5m,作为辅助数据,对植被覆盖度提取结果进行精度检验。它与 TM 影像基本同期,植被处于同一物候期,对植被覆盖度提取结果进行精度检验可信度较高。

2 基本原理和技术流程

2.1 基本原理

LSMM(linear spectral mixture model),被定义为:像元在某一光谱波段的反射率是由构成像元的基本组分的反射率以其所占像元面积比例为权重系数的线性组合^[9]。数学模型如下:

$$R_{i\lambda} = \sum_{k=1}^n f_{ki} C_{k\lambda} + \varepsilon_{i\lambda} \quad (1)$$

$$\sum_{k=1}^n f_{ki} = 1 \quad (k = 1, 2, 3, \dots, n) \quad (2)$$

式中, $R_{i\lambda}$ 为第 λ 波段第 i 像元的光谱反射率; f_{ki} 为对应于 i 像元的第 k 个基本组分所占的分量值; $C_{k\lambda}$ 为第 k 个基本组分在第 λ 波段的光谱反射率; $\varepsilon_{i\lambda}$ 为残差; n 为基本组分数目, m 为可用波段数, 波段数要大于 n ($n \leq m+1$), 以便利用最小二乘法求解。

评价模型用残差 $\varepsilon_{i\lambda}$ 或均方根误差 $RMSE$ 表示:

$$RMSE = \left[\sum_{j=1}^n (\varepsilon_{i\lambda})^2 / n \right]^{1/2} \quad (3)$$

FCM 是一个聚类算法, 可用于遥感监督分类, 通过训练指定各类别的中心, 像元 i 属于类别 j 的模糊隶属度值为 u_{ij} 通过下式计算:

$$u_{ij} = \left[\sum_{k=1}^c \left(\frac{d_{ij}^2}{d_{ik}^2} \right)^{1/(m-1)} \right]^{-1} \quad (4)$$

式中, d_{ij} 是像元 i 到类别 j 中心的距离, d_{ik} 是像元 i 到第 k 个类别中心的距离, c 是类别数, m 是用户定的模糊参数^[10]。

通过 FCM 计算得到的模糊隶属度值表示像元对各训练类别的隶属程度。FCM 算法具有概率约束条件, 它要求计算得到的像元模糊隶属度值的和必须为 1。该算法用于监督分类, 人工选取训练样本需要对研究区有充足的先验知识, 分类结果的好坏与操作人员的知识背景有很大关系。研究中采用交互式端元选择技术, 一方面借助 N 维可视化器, 选择拐角处的纯净像元, 另一方面把选出的像元映射到影像上, 判断其合理性, 反复检验后确定城市土地覆被分类的端元样本, 探寻提高 FCM 提取植被覆盖度专题信息精度的方法。

2.2 技术流程

用两种方式提取海淀区的植被覆盖度, 并对结果进行精度检验, 技术流程如图 2 所示。

处理流程大致包括 3 部分: TM 数据预处理、两种算法提取植被覆盖度以及利用 SPOT5 数据进行精度检验。其中预处理部分主要是辐射校正、几何纠正以及 4 种端元的选取。另外, 在利用 FCM 提取植被覆盖度前, 引入交互式端元选择技术, 训练区由端元所在栅格转换后得到, 改进传统处理流程。

3 数据处理

3.1 遥感数据辐射校正和几何纠正

剔除第 6 波段热红外波段后, 为消除传感器造成的内部误差, 将其余波段的 DN 值转化为表观辐射亮度。另外, 鉴于影像获取时研究区上空晴朗无云, 并且区域大气性质较均一, 选取 6 S 大气校正模型^[11]。获取大气校正参数后, 综合表观辐射亮度图像和大气校正参数建模, 得到反射率图像。

为保证后续精度检验, 将 TM 影像和 SPOT5 影像进行配准, 采样点分布均匀, 误差在一个像元以内。按行政边界对影像进行裁剪, 得到海淀区的影像。

3.2 V-H-L-S 模型与交互式端元选取

组分的确定以 Ridd 在 1995 年提出的用于城市形态研究的模型: V-I-S 模型 (植被-不透水面-土壤模型)

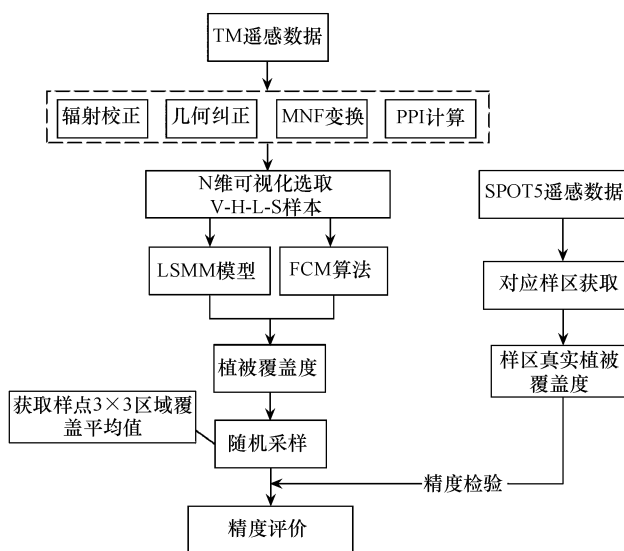


图2 技术流程图

Fig.2 Technique flow chart

为基础。但城市是个复杂的综合体,仅仅基于反射率无法将植被、不透水面、土壤区分开来^[12]。Wu&Murray 在 2003 年结合实际情况,基于 TM 数据,将组分确定为植被、高反射率地物、低反射率地物、土壤 4 类^[13]。本文根据实际情况选用这 4 类,对影像进行一系列处理,依照 V-H-L-S 模型交互式获取端元。

(1) MNF 变换,将主要信息和噪声分离,减少数据的冗余度及波段间的相关性。它由两次主成分变换组成,第 1 次是基于对噪声协方差矩阵的估计,用于分离和重新调节数据中的噪声;第 2 次是对上述处理的图像数据进行一次标准的主成分变换,以提高波谱处理的效果^[14]。经过 MNF 变换生成 6 个波段分量,前 3 个信息含量非常高,占到总的 94.6%,利用前 3 个分量来进行下一步处理。

(2) PPI 计算,是为更好地从混合像元中寻找纯净像元,减少确定端元时所需分析的像元。通常波谱最纯净像元与端元对应^[15]。通过迭代将 N 维散点图映射为一个随机单位向量来计算,其计算结果是一幅 PPI 影像,在该影像上,每个像元的 DN 值与像元被标记为纯净像元的次数对应^[16]。为了进一步减少确定端元时所需分析的像元,文中设定阈值,只选择 DN 值大于 30 的纯净像元。从分布上来看,它们主要分布在西北部的林地、北部的耕地以及北沙河、昆明湖、玉渊潭等水体上。

(3) N 维可视化分析,用于在 N 维空间中选择所需的端元。抽取 MNF 文件的前 3 个分量,并用 PPI 阈值定义的感兴趣区来限定输入 N-维可视化器的像元数量。在 N-维可视化器中,数据云团旋转速度可以控制,处于拐角处的数据云团或与其它数据云团分离得比较远的就是一类端元,结合 N 维可视化器和原始影像,交互地定义这 4 类端元,它们的光谱曲线如图 3 所示。

由图 3 看出,高反射率地物反射率都在 0.2 以上,最高达到 0.7 左右,与其它地物差异明显;低反射率地物在各波段反射率均为 0.1 以下,特征也非常明显;植被的反射率从第 3 波段到第 4 波段有一个反射高峰,符合植被的特征;土壤的反射率基本介于 0.1—0.2 之间。总而言之,各类地物特征明显,具体选取的每类端元的像元数目如表 1 所示。

由表 1 知,植被和低反射率地物这两类端元数目较多,占到了总数的 98% 以上,而由于研究区高反射率地物和土壤纯净像元相对较少,所以选取的端元也较少。

3.3 LSMM 和 FCM 的城市植被覆盖度提取

运用 LSMM 算法以及 FCM 算法,提取城市的植被覆盖度:

(1) LSMM 获取城市植被覆盖度

采用带部分约束条件的最小二乘法,由 LSMM 得到植被覆盖度,且 RMSE 小于 0.1。为方便评价提取精度,对其进行归一化处理。具体结果如图 4a 所示。总体上来说,植被覆盖度空间格局为西北部山区林地高,东南部城区低。

(2) FCM 获取城市植被覆盖度

交互选择得到端元后,通过一系列的转换,将各类端元所在栅格生成训练区。比起纯粹的人工选择,这样更为准确。最后利用 FCM 算法完成植被覆盖度的提取,结果如图 4b 所示。

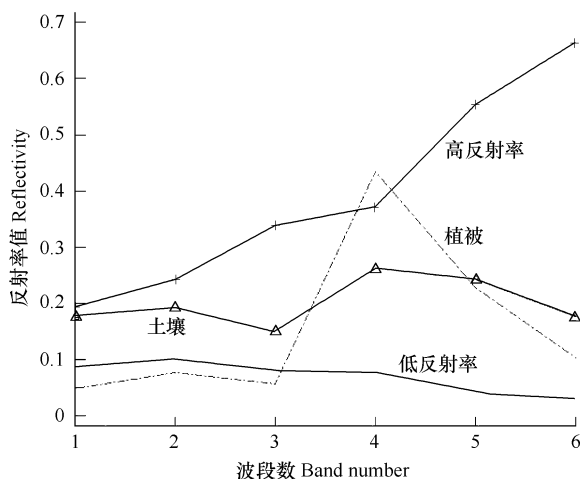


图 3 4 种端元光谱曲线

Fig. 3 Spectral curve of four end-members

表 1 V-H-L-S 端元统计

Table 1 Statistics of V-H-L-S end-members

类别 Category	像元数目 Number of pixels	百分比/% Percentage content
植被 Vegetation	1420	38.27
高反射率 High reflectivity	20	0.54
低反射率 Low reflectivity	2247	60.57
土壤 Soil	23	0.62
总计 Total	3710	100

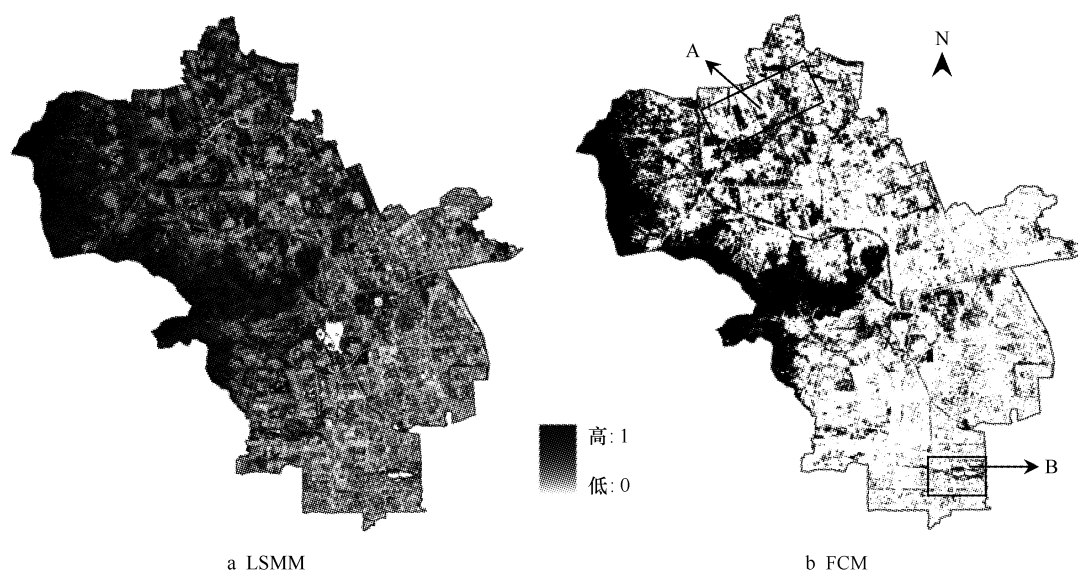


图4 两种方式提取的植被覆盖度
Fig.4 Vegetation coverage extracted by two methods

FCM 算法提取结果在西部及西北部山区林地表现出非常高的植被覆盖,达到 1.0 左右;北部的城乡结合部(图 4b-A)植被覆盖也非常高;城区个别地区,如玉渊潭公园附近(图 4b-B),也体现出较高的植被覆盖。

总的来说,LSMM 提取结果在大部分地区表现为 0.2—0.3 左右的植被覆盖,明显区别于 FCM 提取结果在这些地区植被覆盖度低于 0.1。

3.4 精度检验

采用 SPOT5 作为参考图像,获取其植被覆盖度,作为检验数据和两种方法提取的结果进行相关性分析,计算相关系数 R 。

在算法分解的植被覆盖度图像上随机采样,共选取 60 个点,采样点分布均匀。为尽量减少 TM 影像和 SPOT5 影像之间几何配准带来的误差影响,在每个采样点上,以该采样点对应的像元为中心,选择一个 3×3 的区域(8100m^2),计算这个像元的平均值,得到采样区植被覆盖度。在 SPOT5 影像上找到采样区对应位置,使用人工解译的方法,勾画植被,统计植被面积,用植被面积和采样区面积(8100m^2)相除得到的结果作为植被覆盖度的检验数据。表 2 为 60 个采样区植被覆盖度(FCM 和 LSMM 算法)和 SPOT5 检验数据。

为进一步评价 FCM 和 LSMM 两种分解算法的精度,又采用均方根误差(RMSE)、系统误差(SE)两个指标来反映。

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\tilde{V}_i - V_i)^2}{N}} \quad (5)$$

$$SE = \frac{\sum_{i=1}^N (\tilde{V}_i - V_i)}{N} \quad (6)$$

式中, N 为样本总数, $\tilde{V}_i$ 为 FCM 或 LSMM 分解得到的植被覆盖度, V_i 为检验数据。另外, RMSE 反映样本的总体精度, SE 可体现系统误差的影响。

使用 SPOT5 影像对提取结果进行精度检验,由表 3 发现 LSMM 算法提取的植被覆盖度与验证数据的相关系数 R 为 0.8252;而 FCM 结果 R 为 0.9381,优于 LSMM 结果。另外,利用 FCM 得到的结果, RMSE 和 SE 均小于 LSMM 结果。不过,两种方法的 RMSE 均偏大(均高于 10%)。研究区覆被类型较为复杂,也是导致

RMSE 偏高的原因之一。总的来说,改进后的 FCM 分解精度较高,较好地提取了北京市海淀区 2007 年 5 月植被覆盖度。

表 2 各采样区植被覆盖度和检验数据
Table 2 Vegetation coverage of each sampling area and inspection data

线性光谱分解结果 LSMM-based data			改进的模糊 C-均值分解结果 Improved FCM-based data			SPOT5 检验数据 SPOT5 inspection data		
0.7006	0.2310	0.3324	1.0000	0.0000	0.0000	0.9935	0.1648	0.0000
0.6695	0.1820	0.5594	1.0000	0.0000	0.8205	0.9858	0.1941	0.7623
0.6948	0.3191	0.1597	0.7795	0.0008	0.0000	0.9098	0.2748	0.0670
0.5390	0.1873	0.3488	0.5208	0.0000	0.0700	0.4669	0.0569	0.3463
0.4638	0.6635	0.2476	0.8367	1.0000	0.0000	0.4924	0.8426	0.0336
0.1171	0.6281	0.3519	0.0000	0.7980	0.0000	0.1399	0.5662	0.0000
0.2888	0.4540	0.7759	0.5728	0.4264	1.0000	0.4490	0.1600	0.9733
0.1384	0.3053	0.6493	0.0000	0.0967	1.0000	0.0674	0.1341	0.9802
0.0953	0.3439	0.1270	0.0000	0.0773	0.0000	0.0000	0.0261	0.0000
0.2049	0.3942	0.2222	0.0000	0.3333	0.0000	0.1313	0.2373	0.0649
0.4244	0.2081	0.1471	0.0009	0.0000	0.0000	0.1157	0.0000	0.0000
0.6820	0.2323	0.8460	1.0000	0.0000	1.0000	1.0000	0.1169	0.9848
0.3040	0.2097	0.4233	0.0000	0.0104	1.0000	0.0597	0.2203	0.9580
0.3074	0.1671	0.2803	0.0000	0.0000	0.0000	0.0000	0.0789	0.1447
0.2655	0.2951	0.6337	0.0000	0.0000	0.0000	0.1067	0.4372	0.0519
0.2702	0.1390	0.1609	0.0000	0.0000	0.0000	0.0000	0.0153	0.0542
0.4008	0.2319	0.4877	0.1898	0.1111	0.0652	0.4994	0.2121	0.2287
0.5985	0.4040	0.4435	0.6406	0.1852	0.0000	0.7293	0.3085	0.1624
0.3160	0.3331	0.4416	0.0000	0.0000	0.6143	0.0000	0.2165	0.4476
0.1273	0.5074	0.1058	0.0000	0.2080	0.0000	0.0000	0.2580	0.0000

4 结论与讨论

本文以北京市海淀区植被专题信息提取为例,探寻基于混合像元分解技术估算城市植被覆盖度的方法。利用 LSMM 和 FCM 提取了研究区 2007 年 5 月 Landsat TM 遥感图像的植被分量,获取了城市植被覆盖度专题信息,利用同期 SPOT5 高分辨率遥感图像进行精度验证,结论如下:

(1)采用带部分约束条件的最小二乘法,进行 LSMM 混合像元分解,由于遥感图像上局部噪声、大气效应、环境辐射等非线性因素的影响,植被分量提取难以避免不合理值,表现在植被覆盖度提取结果与检验数据相关系数 R 为 0.8252, RMSE、SE 偏大,城区植被覆盖信息表征略显不足。

(2)为提高训练样本选择的准确度,本文引入城区 V-H-L-S 端元交互式选择技术,改进了传统 FCM 方法的处理流程,通过 MNF、PPI、N 维可视化处理,有效地降低了主观性。另外,在图像中存在模糊性和不确定性时,FCM 算法体现出模糊数学的优势,相比 LSMM 混合像元分解,FCM 算法可以较好地处理其他要素的非线性影响。结果表明,改进的 FCM 结果与高分检验数据相关系数 R 达 0.9381, RMSE、SE 较 LSMM 分解结果小,较好地提取了城区植被覆盖度。

表 3 相关性分析及两种方法误差比较
Table 3 Correlation analysis and error comparison of two methods

项目 Item	相关系数 R	均方根误差 RMSE	系统误差 SE
线性光谱分离结果 LSMM-based data	0.8252	21.34%	6.44%
改进的模糊 C-均值分解结果 improved FCM-based data	0.9381	13.92%	1.94%

References:

- [1] Chen Y H, Li X B, Shi P J. Landscape spatial-temporal pattern analysis on change in the fraction of green vegetation based on remotely sensed data: a case study in Haidian District, Beijing. *Acta Ecologica Sinica*, 2002, 22(10):1581-1586.
- [2] Zhu S L. The classification of remotely-sensed images with mixels. *Journal of PLA Surveying and Mapping Institute*, 1995, 12(4):276-278.
- [3] Ichoku C, Karnieli A. A review of mixture modeling techniques for sub-pixel land cover estimation. *Remote Sensing of Reviews*, 1996, 13(34):161-186.
- [4] Small C. Estimation of urban vegetation abundance by spectral mixture analysis. *International Journal of Remote Sensing*, 2001, 22(7):1305-1334.
- [5] Small C. The Landsat EMT + spectral mixing space. *Remote Sensing of Environment*, 2004, 93(1/2):1-17.
- [6] Daniel C H, Chang C I. Fully constrained least squares linear spectral mixture analysis method for material quantification in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 2001, 39(3):529-545.
- [7] Yang W, Chen J, Matsushita B, Gong P, Chen C X. A new spectral mixture analysis method based on spectral correlation matching. *Journal of Remote Sensing*, 2008, 12(3):454-461.
- [8] Huo D M, Liu G H, Luo J C. Possibilistic c-Means algorithm improving the pixel unmixing of remote sensed image. *Journal of Remote Sensing*, 2005, 9(2):131-137.
- [9] Zhao Y S. The Principle and Method of Analysis of Remote Sensing Application. Beijing: Science Press, 2003: 330-334.
- [10] Bezdek J C, Ehrlich R, Full W. FCM: the fuzzy c-means clustering algorithm. *Computer and Geosciences*, 1984, 10(2/3):191-203.
- [11] Vermote E F, Tanre D, Deuze J L, Herman M, Morcette J J. Second simulation of the satellite signal in the solar spectrum, 6S: an overview. *IEEE Transactions on Geoscience and Remote Sensing*, 1997, 35(3):675-686.
- [12] Powell R L, Roberts D A, Dennison P E, Hess L L. Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazil. *Remote Sensing of Environment*, 2007, 106(2):253-267.
- [13] Wu C S, Murray A T. Estimating impervious surface distribution by spectral mixture analysis. *Remote Sensing of Environment*, 2003, 84(4):493-505.
- [14] Green A A, Berman M, Switzer P, Craig M D. A transformation for ordering multispectral data in terms of image quality with implications for noise removal. *IEEE Transactions on Geoscience and Remote Sensing*, 1988, 26(1):65-74.
- [15] Boardman J W. Automating spectral unmixing of AVIRIS data using convex geometry concepts: In summaries, Forth Annual JPL Airborne Geoscience Workshop. Pasadena, CA, US: JPL Publication, 1993, 26(1):11-14.
- [16] Boardman J W, Kruse F A, Green R O. Mapping target signatures via partial unmixing of AVIRIS data: In summaries, Fifth JPL Airborne Earth Science Workshop. Pasadena, CA, US: JPL Publication, 1995, 1(1):23-26.

参考文献:

- [1] 陈云浩, 李晓兵, 史培军. 基于遥感的植被覆盖变化景观分析——以北京海淀区为例. *生态学报*, 2002, 22(10):1581-1586.
- [2] 朱述龙. 基于混合像元的遥感图像分类技术. *解放军测绘学院学报*, 1995, 12(4):276-278.
- [7] 杨伟, 陈晋, 松下文经, 宫鹏, 陈春晓. 基于相关系数匹配的混合像元分解算法. *遥感学报*, 2008, 12(3):454-461.
- [8] 霍东民, 刘高焕, 骆剑承. 基于 PCM 改进算法的遥感混合像元模拟分析. *遥感学报*, 2005, 9(2):131-137.
- [9] 赵英时. 遥感应用分析原理与方法. 北京: 科学出版社, 2003: 330-334.