

# 鄱阳湖 5 种典型植被高光谱特征波段选择 与光谱分类识别

曾帅, 况润元, 陈彦兵

江西理工大学建筑与测绘工程学院, 江西 赣州 341000

**摘要** 光谱特征波段的选取是植被高光谱分类识别的重要基础之一。利用鄱阳湖 5 种典型植被的实测高光谱数据, 在对数据进行预处理和分析的基础上, 提出了一种基于均值极差阈值法的光谱特征波段选择方法, 并利用马氏距离-光谱角法对不同植被种类进行识别。结果表明: 所提方法有效提取了植被间的光谱特征波段, 分别为 1111~1132 nm、1466~1522 nm 和 1577~1750 nm, 三个波段全部位于红外区域; 在光谱特征波段范围内, 利用马氏距离-光谱角法可对不同植被类型进行有效识别, 其中南荻的光谱分类精度最高, 灰化藁草的光谱分类精度最低, 为 84%, 总体分类精度为 91%, 分类效果较好。

**关键词** 光谱学; 光谱特征波段; 光谱特征; 分类

**中图分类号** O433 **文献标识码** A

**doi:** 10.3788/LOP54.123002

## Hyperspectral Characteristic Band Selection and Spectral Classification of Five Typical Vegetation in Poyang Lake

Zeng Shuai, Kuang Runyuan, Chen Yanbing

*School of Architectural and Surveying & Mapping Engineering, Jiangxi University of  
Science and Technology, Ganzhou, Jiangxi 341000, China*

**Abstract** The selection of spectral characteristic band is one of the important basis of plant hyperspectral classification. On the basis of measured hyperspectral data of five typical vegetation in Poyang Lake and data preprocessing and analysis, a method of spectral characteristic band selection based on the average and range threshold method is proposed, and the Mahalanobis distance-spectral angle method is used to identify the species of different vegetation. The results show that the proposed method effectively extracts the spectral characteristic band of the vegetation, the band is 1111-1132 nm, 1466-1522 nm, and 1577-1750 nm, respectively, and all of them are located in the infrared region. In the spectral characteristic band, the Mahalanobis distance-spectral angle method can effectively identify different vegetation types, the spectral classification accuracy of *Triarrhena* is the highest, the accuracy of *Cynodon* is 84%, and the overall classification accuracy is 91%, which shows that the classification effect is good.

**Key words** spectroscopy; spectral characteristic band; spectral characteristics; classification

**OCIS codes** 300.6550; 110.4234; 350.5610

## 1 引言

高光谱数据具有波段多、光谱分辨率高等特点, 为地物光谱分类识别研究提供了极大便利<sup>[1-3]</sup>, 但它同时也面临着数据量大、谱间相关性高及数据冗余严重等问题。通常情况下, 某种地物只在某些波长范围内才与

收稿日期: 2017-06-29; 收到修改稿日期: 2017-07-14

基金项目: 国家自然科学基金(41101322)、江西省教育厅科学技术研究项目(GJJ160617)

作者简介: 曾帅(1993—), 男, 硕士研究生, 主要从事高光谱遥感方面的研究。E-mail: jydlzs@163.com

导师简介: 况润元(1976—), 男, 博士, 副教授, 硕士生导师, 主要从事遥感机理方面的研究。

E-mail: rykuang@163.com(通信联系人)

其他地物具有显著的光谱差异特征,若要对所有波长范围内的数据进行分析处理,势必会增加额外的工作量。故利用高光谱数据对不同植被种类识别时,通过降低数据维度提取出特征波段显得尤为重要<sup>[4]</sup>。此外,光谱相似性测度是高光谱数据分类的基础,基于地物间的光谱特征波段,运用高效的相似性测度方法有助于提高不同地物间的光谱分类识别精度。

在光谱特征波段提取方法的研究方面,陈永刚等<sup>[5]</sup>提出了一种可用于提取不同树种光谱特征波段的均值置信区间带法,并利用曼哈顿距离和最小-最大区间相似度法对光谱识别效果进行了验证;刘雪华等<sup>[6]</sup>采用马氏距离法和主成分分析法提取了北京地区常见湿地植被的光谱特征波段,并构建了有效识别不同植被的判别模型;Dian 等<sup>[7]</sup>以中国中部 13 种常见树种为研究对象,对比分析了不同分类方法的分类效果。以上学者在光谱特征波段提取方法和植被识别研究方面都取得了一定成果,但目前尚未见到利用统计学中的极差法对不同植被光谱特征波段进行提取及光谱识别的研究报道。在光谱曲线测度和分类识别方法的研究方面,Kumar 等<sup>[8]</sup>提出了利用相关系数和光谱信息离散结合的相似性测度方法;Kong 等<sup>[9]</sup>提出了一种结合多种光谱特征的光谱泛相似测度方法;张浚哲等<sup>[10]</sup>在对欧式距离和相关系数分析的基础上,提出了欧式距离-相关数据的测度方法。虽然以上研究都取得了较好的效果,但在方法上仍有进一步研究的空间。

基于以上几点,本文以鄱阳湖 5 种典型湿地植被实测光谱数据为研究对象,提出了一种基于均值极差阈值的植被光谱特征波段筛选方法,并在该方法提取出的光谱特征波段基础上,利用马氏距离-光谱角相似性测度法对不同种类典型植被进行分类识别,最后统计出不同典型植被的光谱分类精度,并对植被分类精度进行分析评价。

## 2 方法及原理

### 2.1 数据的获取与预处理

2016 年 11 月,本课题组利用美国 ASD FieldSpec 4 便携式地物光谱仪采集鄱阳湖湿地 5 种典型植被——狗牙根、南荻、水蓼、灰化藁草、芦苇——的光谱数据。光谱采样波长范围是 350~2500 nm,对于波长范围分别为 350~1000 nm 和 1000~2500 nm 的光谱采样间隔分别为 1.4 nm 和 2 nm,光纤前视场角为 25°。在光谱测量过程中,选取植被单一、长势良好、植被覆盖度较高的植丛布置样方,样方大小设置为 1 m×1 m。为避免样方之间相互影响,样方间的距离大于 3 m。为使数据具有代表性,采样点均匀分布于研究区内。为保证良好的光照条件,测量时间选定在 10:00—14:30。采用垂直观测法在距植被冠层上方 1 m 左右处测取数据,每个采样点测取 10 条光谱曲线,求取均值光谱作为最终的样点数据。测量过程中最多每隔 20 min 进行一次白板优化。植被光谱数据采集后,需要对数据进行预处理,包括剔除异常光谱数据和去除噪声波段。本课题组参照王锦地等<sup>[11]</sup>提到的强度大小检验法剔除异常光谱数据。剔除异常光谱数据后,再利用 ViewSpecPro 软件对噪声波段进行去除处理,保留的有效波段范围为 400~1350 nm、1450~1750 nm 和 2000~2400 nm。数据预处理后,随机抽取每种植被的 10 个样点数据用以实验分析,另取每种植被的 50 个样点数据对提出的方法进行分类精度检验。

### 2.2 均值极差阈值法

高光谱数据为地物光谱分类识别研究提供了极大的便利,但同时也面临着数据量大、谱间相关性高等问题,通过高光谱降维技术可以有效降低数据的冗余度,保留有效信息。高光谱数据降维主要包括基于变换的特征提取和基于非变换的特征选择两种方法<sup>[12]</sup>。相对于前者,基于非变换的特征选择法的优点主要是,在不改变光谱实际物理意义的前提下筛选出可以有效识别植被的特征波段,在操作上更加方便快捷。该方法的原理为通过简化原始光谱特征空间,筛选出一个包含目标地物主要特征光谱信息的子集波段,用较少的特征完成数据的收集和分析处理,同时降低了数据维度。

在统计学中,极差用以反映一组数据之间的最大差距。在光谱测量过程中,由于外界因素的影响,实测光谱数值会在一定范围内波动,这个波段范围的大小可以用极差来表示。本方法的描述为:假定在光谱波段  $i$  处,两种植被 a、b 的光谱数据均值差的绝对值大于极差之和,则可认为在波段  $i$  处两种植被存在显著差异性,即波段  $i$  为两种植被可相互区别的光谱特征波段。与前人所提方法不同的是,本方法能捕捉到植被光谱间反射率值的微小差异,在保留植被光谱物理意义的同时,能提取出植被光谱的差异波段,在操作上更加方

便快捷。此外,为方便后期实测高光谱数据与高光谱遥感影像结合,规定选出的光谱差异波段范围必须大于高光谱遥感数据的光谱分辨率(10 nm 左右)。相关的计算公式为

$$|\overline{X_{a(i)}} - \overline{X_{b(i)}}| > R_{a(i)} + R_{b(i)}, \quad (1)$$

$$R_{a(i)} = S_{a(i), \max} - S_{a(i), \min}, \quad (2)$$

$$R_{b(i)} = S_{b(i), \max} - S_{b(i), \min}, \quad (3)$$

式中 $\overline{X_{a(i)}}$ 、 $\overline{X_{b(i)}}$ 分别为植被 a、b 在波段  $i$  处的均值光谱值, $R_{a(i)}$ 、 $R_{b(i)}$ 分别为植被 a、b 在波段  $i$  处的光谱极差值, $S_{a(i), \max}$ 、 $S_{b(i), \max}$ 分别为植被 a、b 在波段  $i$  处的光谱最大值, $S_{a(i), \min}$ 、 $S_{b(i), \min}$ 分别为植被 a、b 在波段  $i$  处的光谱最小值。

### 2.3 马氏距离光谱角(MD\_SA)法

光谱曲线具有形状和幅度两个特征<sup>[13]</sup>,不同地物光谱曲线形状和幅度的差异使得地物之间的分类识别成为可能。利用光谱曲线的幅度特征进行识别分类的方法主要有欧式距离法<sup>[14]</sup>、马氏距离法<sup>[15]</sup>等,利用光谱曲线的形状特征进行分类的方法主要有光谱角余弦法<sup>[16]</sup>、相关系数法<sup>[17]</sup>等。已有的研究成果表明<sup>[8-10]</sup>,利用光谱单一特征无法全面反映地物光谱间的相似性和差异性,结合多种相似性测度方法的光谱曲线分类精度优于单一相似性测度方法。相对于其他地域类型植被,鄱阳湖湿地植被受湖区水位影响较大,植被光谱具有高度相似性,植被种类间更难分类识别。为此,将马氏距离与光谱角两种不同特征测度方法结合起来构建了一种新的光谱相似度测度分类方法。

马氏距离用以表示数据的协方差距离,马氏距离法是一种计算两个未知样本集之间相似度的有效方法。与其他距离法相比,马氏距离既消除了变量间的相关性干扰,又不受量纲的影响。光谱角法是一种基于标准光谱与样本光谱相似性对光谱曲线进行识别分类的光谱匹配方法,其原理是将光谱作为矢量映射到多维空间,根据公式计算出某一样本光谱与标准光谱之间夹角的余弦值,从而确定两者的匹配程度<sup>[18]</sup>。在新方法中,马氏距离法可以计算不同光谱曲线在幅度特征上的细微差别,而光谱角法可以判别光谱曲线的整体形状,即新方法同时兼顾了形状和幅度两个特征。为了便于后续进行光谱分类识别分析,设定方法遵循“值越小,光谱越相似”的距离聚类原则。根据此原则,需要对光谱角值进行翻转(即 1-光谱角),其具体公式表述为

$$L_{MD\_SA(x,y,i)} = L_{MD,(x,y,i)} \cdot L_{SA,(x,y,i)}, \quad (4)$$

$$L_{MD,(x,y,i)} = (\mathbf{x}_i - \mathbf{y}_i)^T \mathbf{S}^{-1} (\mathbf{x}_i - \mathbf{y}_i), \quad (5)$$

$$L_{SA,(x,y,i)} = 1 - \arccos \left( \frac{\sum_{i=1}^n \mathbf{x}_i \mathbf{y}_i}{\sqrt{\sum_{i=1}^n \mathbf{x}_i^2} \sqrt{\sum_{i=1}^n \mathbf{y}_i^2}} \right), \quad (6)$$

式中 $L_{MD,(x,y,i)}$ 为不同光谱曲线  $x$ 、 $y$  在  $i$  波段处的马氏距离, $\mathbf{x}_i$ 、 $\mathbf{y}_i$ 分别为光谱曲线  $x$ 、 $y$  在  $i$  波段处的光谱值, $\mathbf{S}^{-1}$ 为协方差矩阵的逆矩阵, $L_{SA,(x,y,i)}$ 为光谱曲线  $x$ 、 $y$  在  $i$  波段处的夹角余弦值。植被间 MD\_SA 值的计算主要在统计产品与服务解决方案(SPSS)软件和 Excel 中共同完成。在后续的光谱曲线分类识别中,光谱曲线  $x$ 、 $y$  分别设定为待检验光谱和参考光谱,其中参考光谱为每种植被 10 条实验光谱的均值光谱,另取每种植被 50 条待分类验证光谱进行分析。本方法是一种用“距离”评价个体间相似程度,根据距离最近原则判定并归并到同一个类别中的方法。

## 3 实验与分析

### 3.1 光谱曲线分析及特征波段提取

图 1 为每种植被的 10 条反射光谱曲线,5 种植被曲线都表现出了植被特有的曲线特征,5 种植被光谱曲线具有高度的相似性。受植被本身生化组分和冠层结构的影响,植被光谱曲线在幅度特征上表现出了细微的差别,通过分析不同植被间的细小差异,提取出有效识别不同植被的光谱差异波段,然后利用相关的测度方法,对不同植被种类光谱曲线进行分类识别。在可见光波段,5 种植被的光谱曲线都表现为低反射率,在蓝光波段(400~530 nm)和红光波段(550~730 nm)的吸收较为强烈,形成了蓝谷和红光 2 个吸收谷,介于两者间的绿光波段则形成了 1 个凸起的反射峰。在 670~740 nm 波段的“红边”反射率迅速增高,增长率为

40%~50%，在 780~950 nm 波段范围内继续保持高反射率值。在近红外波段，以 960 nm 和 1200 nm 为中心形成了水分的吸收带，即吸收谷。图 2 为根据均值极差阈值法提取出的光谱特征波段，5 种植被光谱曲线的光谱特征波段为 1111~1132 nm、1466~1522 nm 和 1577~1750 nm，即 5 种植被光谱曲线在这三个光谱特征波段范围内存在差异。三个光谱特征波段全部位于植被光谱的红外波段，其中 1111~1132 nm 波段范围较窄，1466~1522 nm 和 1577~1750 nm 波段相对较宽。综上所述，使用均值极差阈值法可有效提取出不同光谱曲线的光谱特征波段。

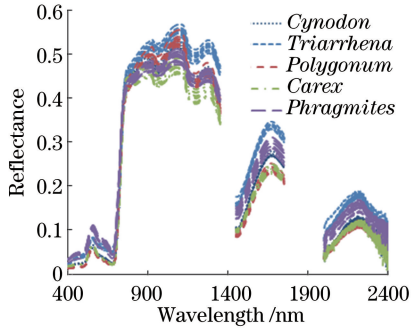


图 1 典型植被的光谱曲线

Fig. 1 Spectral curves of typical vegetation species

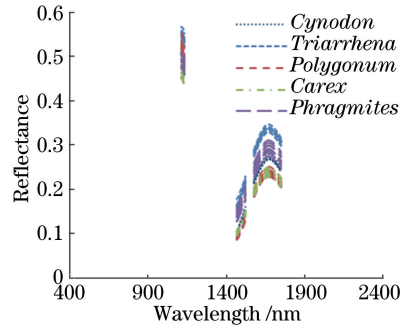


图 2 提取出的光谱特征波段

Fig. 2 Extracted spectral feature band

### 3.2 光谱波段分类识别

基于均值极差阈值法提取出 5 种典型植被的光谱特征波段，利用 MD\_SA 法对每种植被各 50 条验证光谱进行判别分类。首先计算每种植被每条待测样本光谱与参考样本光谱间的 MD\_SA 值，根据“值越小，光谱越相似”的距离聚类原则，将待测样本归为最小 MD\_SA 值对应的参考光谱类别中，最后统计正确/错误分类的待测样本数目，并统计分类精度。表 1 为 5 种典型植被光谱分类结果的混淆矩阵，表中用总分类精度、生产者精度和用户精度等指标来表述最后的光谱识别分类效果。其中生产者精度表示本类样本划分到本类的比例，用户精度表示分类结果中确实属于该类样本的比例，总分类精度表示正确分类的样本占总参与分类样本的比例。

从表 1 中可以看出，基于均值极差阈值法提取出的光谱特征波段，运用 MD\_SA 法，根据“值越小，光谱越相似”原则对 5 种典型植被待分类光谱进行分类识别，总体精度达到了 91%，分类识别效果良好。在 5 种典型植被中，南荻的正确分类精度最高，达到了 100%，表明其在光谱特征波段上与其他植被能较好地区分开来；灰化藁草的正确分类精度最低，为 84%，这是由于其与南荻、水蓼在 1466~1522 nm 和 1577~1750 nm 两个光谱特征波段上反射率值接近，容易造成错误分类；同样，水蓼与芦苇在 1111~1132 nm 波段范围内容易误分，水蓼与灰化藁草在 1466~1522 nm 波段范围内容易错分；狗牙根的光谱主要易与南荻和芦苇的光谱混淆，芦苇的光谱主要易与水蓼的光谱混淆。混淆矩阵分析表明，综合利用均值极差阈值法和 MD\_SA 法，根据最小聚类原则，可对鄱阳湖 5 种典型植被光谱进行有效的分类识别，分类效果较好，总体分类精度可达 91%。

表 1 分类混淆矩阵

Table 1 Confusion matrix of classification

Vegetation	<i>Cynodon</i>	<i>Triarrhena</i>	<i>Polygonum</i>	<i>Carex</i>	<i>Phragmites</i>	Producer accuracy /%
<i>Cynodon</i>	45	2	0	0	3	90
<i>Triarrhena</i>	0	50	0	0	0	100
<i>Polygonum</i>	0	0	43	5	2	86
<i>Carex</i>	0	3	5	42	0	84
<i>Phragmites</i>	0	0	3	0	47	94
User accuracy /%	100	91	90	89	90	91*

Note: \* represents total classification accuracy.

## 4 结 论

针对淡水湖泊湿地植被光谱相似度高、难以分类识别等特点,基于鄱阳湖 5 种植被实测高光谱数据,引入统计学中的极差概念,提出了一种基于均值极差阈值法的光谱特征波段选择方法,利用该方法提取出 5 种植被的光谱特征波段,运用 MD\_SA 法,根据最小聚类原则对 5 种典型植被进行光谱分类识别,并统计最后的分类精度,得出以下结论:

1) 提出的方法通过剔除植被间的光谱重叠区域,有效提取出了不同植被间的光谱特征波段,同时有效降低了数据的维度,提取出的光谱特征波段分别为 1111~1132 nm、1466~1522 nm 和 1577~1750 nm,这些光谱特征波段全部位于红外波段,而红外波段正是植被对自然光的高反射平台区,不同植被在此区间的光谱反射率值差异较为显著。该方法在没有改变光谱物理意义的条件下提取出了植被光谱特征波段,在操作上更加方便快捷。

2) 光谱曲线具有形状和幅度两个特征。传统的单种相似性测度方法不能同时兼顾形状和幅度两个特征,在光谱相似性高的识别分类应用中往往具有较低的精度。结合马氏距离和光谱角法提出了新的光谱相似性测度方法——MD\_SA 法,通过计算待检验光谱与参考光谱间的 MD\_SA 值,依据最小聚类原则判别不同植被的种类归属,最后统计不同植被光谱的分类精度。南荻的光谱分类精度可达 100%,灰化藁草的分类精度最低,为 84%,总体分类精度为 91%,总体分类效果较好。

地物光谱识别是高光谱遥感地物属性识别分类的基础,不同类型地物可根据某些光谱差异显著的波段进行有效识别,如何对这些光谱差异波段范围进行界定和筛选则需要一定的方法,而所提方法解决了这一问题。对植被光谱特征波段进行选取,有利于卫星传感器波段的选择,可以更好地应用于多光谱或高光谱影像数据的分类研究中。研究成果可为湿地遥感影像制图提供参考。

## 参 考 文 献

- [1] Yu Yang, Zhao Nanjing, Wang Yin, *et al.* Research on the characteristics of lead contained in lead slime by laser-induced breakdown spectroscopy[J]. Chinese J Lasers, 2014, 41(7): 0715001.  
余洋, 赵南京, 王寅, 等. 含铅污泥中铅的激光诱导击穿光谱特征研究[J]. 中国激光, 2014, 41(7): 0715001.
- [2] Guo Yulong, Li Yunmei, Wang Qiao, *et al.* A new water chlorophyll-a index and analysis of its application potential [J]. Acta Optica Sinica, 2015, 35(11): 1101003.  
郭宇龙, 李云梅, 王桥, 等. 一种新的水体叶绿素 a 指数及其应用潜力分析[J]. 光学学报, 2015, 35(11): 1101003.
- [3] Zhang Dong, Tashpolat · Tiyp, Zhang Fei, *et al.* Effect of fractional differential algorithm on hyperspectral data of saline soil[J]. Acta Optica Sinica, 2016, 36(3): 0330002.  
张东, 塔西普拉提·特依拜, 张飞, 等. 分数阶微分算法对盐渍土高光谱数据的影响研究[J]. 光学学报, 2016, 36(3): 0330002.
- [4] Tong Q X, Xue Y Q, Zhang L F. Process in hyperspectral remote sensing science and technology in China over the past three decades[J]. IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing, 2014, 7(1): 70-91.
- [5] Chen Yonggang, Ding Lixia, Ge Hongli, *et al.* Hyperspectral feature band selection based on mean confidence interval and tree species discrimination[J]. Spectroscopy and Spectral Analysis, 2011, 31(9): 2462-2466.  
陈永刚, 丁丽霞, 葛宏立, 等. 基于均值置信区间带的高光谱特征波段与树种识别[J]. 光谱学与光谱分析, 2011, 31(9): 2462-2466.
- [6] Liu Xuehua, Sun Yan, Wu Yan. Reduction of hyperspectral dimensions and construction of discriminating models for identifying wetland plant species[J]. Spectroscopy and Spectral Analysis, 2012, 32(2): 459-464.  
刘雪华, 孙岩, 吴燕. 光谱信息降维及判别模型建立用于识别湿地植物物种[J]. 光谱学与光谱分析, 2012, 32(2): 459-464.
- [7] Dian Y Y, Fang S H, Le Y, *et al.* Comparison of the difference classifiers in vegetation species discrimination using hyperspectral reflectance data[J]. Journal of the Indian Society of Remote Sensing, 2014, 42(1): 61-72.
- [8] Kumar M N, Seshasai M V R, Prasad K S V, *et al.* A new hybrid spectral similarity measure for discrimination among *Vigna* species[J]. International Journal of Remote Sensing, 2011, 32(14): 4041-4053.

- [9] Kong X B, Shu N, Tao J B, *et al.* A new spectral similarity measure based on multiple features integration[J]. *Spectroscopy and Spectral Analysis*, 2011, 31(8): 2166-2170.
- [10] Zhang Junzhe, Zhu Wenquan, Zheng Zhoutao, *et al.* Comparative study on similarity measures of hyperspectral remote sensing data[J]. *Science of Surveying and Mapping*, 2013, 38(6): 33-36.  
张浚哲, 朱文泉, 郑周涛, 等. 高光谱数据的相似性测度对比研究[J]. *测绘科学*, 2013, 38(6): 33-36.
- [11] 王锦地, 张立新, 柳钦火, 等. 中国典型地物波谱知识库[M]. 北京: 科学出版社, 2009: 71-73.
- [12] 童庆禧, 张兵, 郑兰芬. 高光谱遥感——原理、技术与应用[M]. 北京: 高等教育出版社, 2006: 129-136.
- [13] Dundar M M, Landgrebe D A. Toward an optimal supervised classifier for the analysis of hyperspectral data[J]. *IEEE Transaction on Geoscience and Remote Sensing*, 2004, 42(1): 271-277.
- [14] Jain A K, Duin R, Mao J C. Statistical pattern recognition: a review[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000, 22(1): 4-37.
- [15] Lin Haijun, Zhang Huifen, Gao Yaqi, *et al.* Mahalanobis distance based hyperspectral characteristic discrimination of leaves of different desert tree species[J]. *Spectroscopy and Spectral Analysis*, 2014, 34(12): 3358-3362.  
林海军, 张绘芬, 高亚琪, 等. 基于马氏距离的荒漠树种高光谱识别[J]. *光谱学与光谱分析*, 2014, 34(12): 3358-3362.
- [16] Kruse F A, Lefkoff A B, Boardman J W, *et al.* The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data[J]. *Remote Sensing of Environment*, 1993, 44(2/3): 145-163.
- [17] Vandermeer F, Bakker W. Cross correlogram spectral matching: application to surface mineralogical mapping by using AVIRIS data from Cuprite, Nevada[J]. *Remote Sensing of Environment*, 1997, 61(3): 371-382.
- [18] Sun Yanli, Zhang Xia, Shuai Tong, *et al.* Radiometric normalization of hyperspectral satellite images with spectral angle distance and Euclidean distance[J]. *Journal of Remote Sensing*, 2015, 19(4): 618-626.  
孙艳丽, 张霞, 帅通, 等. 光谱角-欧式距离的高光谱图像辐射归一化[J]. *遥感学报*, 2015, 19(4): 618-626.