

# 考虑空间相关性的半监督局部保持投影的高光谱图像特征提取

黄冬梅<sup>1,2</sup>, 张晓桐<sup>1</sup>, 张明华<sup>1\*</sup>, 宋巍<sup>1</sup>, 王龔<sup>1</sup>

<sup>1</sup>上海海洋大学信息学院, 上海 201306;

<sup>2</sup>上海电力大学, 上海 200090

**摘要** 由于高光谱图像具有波段多、波段间信息冗余、空间信息相关等特点, 提出一种考虑空间相关性的半监督局部保持投影(LPP)的高光谱图像特征提取算法(LPP-SCSSFE)。该算法利用保存高光谱图像空间近邻结构的空间距离, 及保持图像光谱相似性的类内判别权值和类间判别权值, 定义新的同物异谱、同谱异物像元权值计算函数, 结合 LPP 提取高光谱图像特征, 从而最大化同类地物间的相似性和异类地物间的差异性。在 Indian Pines 和 Pavia University 两个数据集上, 通过高光谱图像分类实验对本文提出的 LPP-SCSSFE 算法进行验证, 算法最高总体分类精度分别达到 87.50% 和 91.29%, 优于现有的特征提取算法。结果表明, 本文算法充分考虑高光谱图像的空间相关性和光谱相似性, 能够有效提取出有代表性的特征, 提高分类精度。

**关键词** 图像处理; 特征提取; 局部保持投影; 高光谱图像; 半监督; 空间相关性

中图分类号 O436

文献标识码 A

doi: 10.3788/LOP56.021003

## Feature Extraction of Hyperspectral Images Based on Semi-Supervised Locality Preserving Projection with Spatial-Correlation

Huang Dongmei<sup>1,2</sup>, Zhang Xiaotong<sup>1</sup>, Zhang Minghua<sup>1\*</sup>, Song Wei<sup>1</sup>, Wang Yan<sup>1</sup>

<sup>1</sup> College of Information Technology, Shanghai Ocean University, Shanghai 201306, China;

<sup>2</sup> Shanghai University of Electric Power, Shanghai 200090, China

**Abstract** Based on the characteristics of multi-band, inter-band information redundancy and spatial information correlation of hyperspectral images, a spatially-correlated and semi-supervised feature extraction (SCSSFE) algorithm with locality preserving projection (LPP) is proposed. This algorithm defines a new pixel weight calculation function for the different spectral characteristics with the same objects and the different objects with the same spectral characteristics to preserve the spatial distance and the spectral similarity of hyperspectral image by means of the neighbor structure in image space and the intra-class and inter-class discriminant weights. Then, the features of hyperspectral images are extracted by the weight function combined with LPP. Thus the similarity among the same objects and the discrepancy among the different objects are maximized. The proposed LPP-SCSSFE algorithm is verified through the hyperspectral image classification experiments on the two datasets of Indian Pines and Pavia University. The highest overall classification accuracies of the LPP-SCSSFE algorithm reach 87.50% and 91.29% for the respective datasets, better than those of the existing feature extraction algorithms. These results indicate that the spatial correlation and the spectral similarity of hyperspectral images are fully taken into account in the proposed algorithm, and thus the more representative features are extracted and the classification accuracy is enhanced.

**Key words** image processing; feature extraction; locality preserving; hyperspectral images; semi-supervised; spatial correlation

**OCIS codes** 100.4145; 100.2960; 100.2000

收稿日期: 2018-05-21; 修回日期: 2018-06-01; 录用日期: 2018-07-30

基金项目: 国家自然科学基金(41671431)、上海市科学技术委员会科研计划项目(15590501900)、上海市高校特聘教授(东方学者)项目(TP201638)

\* E-mail: mhzhang@shou.edu.cn

# 1 引言

随着高光谱传感器技术的发展,高光谱遥感图像的应用越来越广泛,如目标检测和土地覆盖调查<sup>[1-2]</sup>等。但高光谱遥感图像的波段通常连续且多达十个甚至数百个<sup>[3]</sup>,且波段间的相关性和冗余度高<sup>[4-5]</sup>,因此对高光谱图像进行处理时计算量大、处理难度高,而且高光谱图像占据了很大的存储空间<sup>[6]</sup>。而对高光谱图像特征进行提取,不仅可以保存图像的有用信息,提高图像的分类精度,而且可以很大程度上减少波段数量,提高处理效率并节省空间存储的成本<sup>[7-8]</sup>。

高光谱特征提取<sup>[9-11]</sup>是指对原始图像的光谱特征进行特征转换,提取出更能提高当前分类效果的新特征。考虑数据是否引入类标<sup>[12]</sup>,现有的特征提取方法可以分为三类,分别为有监督特征提取、无监督特征提取和半监督特征提取<sup>[13]</sup>。无监督特征提取算法包括主成分分析(PCA)<sup>[14]</sup>和局部保持投影(LPP)<sup>[15-17]</sup>;有监督特征提取算法包括局部敏感判别分析(LSDA)<sup>[18]</sup>和邻域保持嵌入(NPE)<sup>[19]</sup>。在高光谱图像领域,为了提高分类精度,有监督特征提取算法引入了大量类标,但是高光谱图像类标信息的成本高且不易获取,而无监督特征提取能挖掘出无类标样本中隐藏的信息,因此,基于半监督思想的特征提取方法成为研究热点<sup>[20]</sup>。

LPP算法<sup>[15]</sup>是He等提出的一种经典的特征提取算法,被应用于人脸识别问题中。该算法的最大优势在于能有效保持图像数据的局部几何特征,成为目前的研究热点<sup>[21]</sup>,但其是一种线性无监督特征提取算法。潘银松等<sup>[22]</sup>在LPP的基础上,考虑了少量的像元类标信息,提出了半监督局部保持投影(SSLPP)算法,并将该算法应用于高光谱图像处理领域;Du等<sup>[23]</sup>提出了基于成对约束的半监督特征提取(SSDR-PCP)算法,相比SSLPP算法,该算法分别考虑了同类像元间和不同类像元间的光谱相似性。以上特征提取方法主要是基于

光谱相似性,但是光谱值相近的两个像元可能在空间分布上距离较大;或者由于光线、拍摄角度不同,同一类地物的两个像元可能呈现的光谱并不相似,故仅基于光谱相似性的特征提取会导致分类结果不够准确。

在实际应用中,高光谱图像既存在光谱信息又存在特有的空间信息,且高光谱图像存在同物异谱、同谱异物的现象,所以仅使用光谱信息不能全面有效地表示高光谱图像的本质特征。因此,本文提出一种考虑空间相关性的半监督LPP高光谱图像特征提取(LPP-SCSSFE)算法,该算法充分利用LPP算法的局部保持特性,在同物异谱、同谱异物像元间权重关系中考虑了光谱相似性和空间相关性,其中包括:1)光谱相似的同类像元的类内判别相似性和空间相关性;2)光谱相似的异类像元的类内判别相似性和空间相关性;3)光谱不相似的像元间的空间相关性及光谱差异性。采用两个公共数据集Indian Pines和Pavia University,通过实验证明,该算法能够最大化同类地物相似性以及不同类地物的差异性,更好地挖掘出高光谱图像隐藏的空间特征。由于充分利用了图像的光谱信息和空间信息,从而进一步提高了图像的分类精度。

## 2 局部保持投影

给定一个在高维空间 $\mathbf{R}^D$ 中图像的像元数据集 $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ ,LPP的目的是得到最优的转换矩阵 $\mathbf{A}=[\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_d]$ ,将 $N$ 个像元映射到低维的子空间 $\mathbf{R}^d$ ( $d \ll D$ )的像元集为 $\{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N\}$ ,其中 $\mathbf{y}_i = \mathbf{A}^T \times \mathbf{x}_i$ 。LPP的光谱相似权重为

$$W_{ij} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / \sigma^2), & \mathbf{x}_j \in \mathbf{J}_k(\mathbf{x}_i) \\ 0, & \mathbf{x}_j \notin \mathbf{J}_k(\mathbf{x}_i) \end{cases}, \quad (1)$$

式中: $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{iD}\}$ ;  $\mathbf{x}_j = \{x_{j1}, x_{j2}, \dots, x_{jD}\}$ ;  $\mathbf{J}_k(\mathbf{x}_i)$ 为 $\mathbf{x}_i$ 的光谱域; $\sigma$ 为调节参数,其取值为所有像元间光谱欧氏距离平均值,表达式为

$$\sigma = \frac{1}{N} \times \sum_{i,j=1}^N \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \dots + (x_{iD} - x_{jD})^2}. \quad (2)$$

LPP的目标函数为

$$\min_{\mathbf{A}} \frac{1}{2} \sum_{i,j} (\mathbf{y}_i - \mathbf{y}_j)^2 W_{ij} = \min_{\mathbf{A}} \frac{1}{2} \sum_{i,j} (\mathbf{A}^T \mathbf{x}_i - \mathbf{A}^T \mathbf{x}_j)^2 W_{ij} = \min_{\mathbf{A}} \mathbf{A}^T \mathbf{X} (\mathbf{D} - \mathbf{W}) \mathbf{X}^T \mathbf{A} = \min_{\mathbf{A}} \mathbf{A}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{A}, \quad (3)$$

式中: $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ ;  $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_N)$ ;  $\mathbf{W}$

为由  $W_{ij}$  元素构成的权值矩阵;  $\mathbf{D}$  为  $\mathbf{W}$  的对角矩阵,  $D_{ii} = \sum_j W_{ij}$ ;  $\mathbf{L} = \mathbf{D} - \mathbf{W}$ , 为拉普拉斯矩阵。引入约束条件  $\mathbf{YDY}^T = \mathbf{A}^T \mathbf{XDX}^T \mathbf{A} = \mathbf{I}$ , 其中  $\mathbf{I}$  为单位矩阵, 则(3)式最优化问题可以转化为求解如下广义特征值问题:

$$\mathbf{XLX}^T \mathbf{A} = \lambda \mathbf{XDX}^T \mathbf{A}. \quad (4)$$

令  $\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_d$  分别为(4)式的前  $d$  个最小非零特征值对应的特征向量, 则有

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_d]. \quad (5)$$

LPP 算法具有局部保持光谱信息的优点, 但是它是无监督的, 且没有考虑像元间的空间关系。

### 3 考虑空间相关性的半监督 LPP 特征提取

本研究考虑了像元的类标信息并结合了图像的光谱相似性和空间相关性, 提出一种考虑空间相关性的半监督 LPP 高光谱图像特征提取算法。

#### 3.1 基于空间信息和光谱信息构建半监督权值

设高维空间  $\mathbf{R}^D$  中存在  $l + u$  个像元数据集  $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l, \mathbf{x}_{l+1}, \dots, \mathbf{x}_{l+u}\}$ ,  $l + u = N$ , 其中前  $l$  个为有类标像元, 类别标签为  $c$ , 各类像元数为  $N_i$ ,  $i = 1, 2, \dots, c$ ; 后  $u$  个是无类标像元。利用像元的类标信息、光谱相似性以及空间相关性, 构建本文的同

物异谱、同谱异物的半监督权重值, 即

$$q_{ij} = \begin{cases} D_{ij}^2 (1 + W_{ij}), & \mathbf{x}_i \in \mathbf{X}_l, \mathbf{x}_j \in \mathbf{X}_l, \\ & \mathbf{x}_j \in \mathbf{J}_k(\mathbf{x}_i), c_i = c_j \\ D_{ij}^2 (1 - W_{ij}), & \mathbf{x}_i \in \mathbf{X}_l, \mathbf{x}_j \in \mathbf{X}_l, \\ & \mathbf{x}_j \in \mathbf{J}_k(\mathbf{x}_i), c_i \neq c_j \\ D_{ij}^2 d, & \mathbf{x}_j \notin \mathbf{J}_k(\mathbf{x}_i) \\ 0, & \text{else} \end{cases}, \quad (6)$$

式中:  $q_{ij}$  为半监督权重矩阵  $\mathbf{Q}$  的元素;  $D_{ij}$  为像元间的空间距离, 即两个像元间坐标的相对位置; 像元  $\mathbf{x}_i$  的空间坐标为  $(p, q)$ , 像元  $\mathbf{x}_j$  的空间坐标为  $(m, n)$ , 则  $D_{ij} = \sqrt{(p-m)^2 + (q-n)^2}$ ;  $d$  为光谱欧氏距离,  $d = |\mathbf{x}_i - \mathbf{x}_j|^2$ 。  $W_{ij} = \exp(-|\mathbf{x}_i - \mathbf{x}_j|^2 / \sigma^2)$ ,  $W_{ij}$  为局部权重值, 令  $(1 + W_{ij})$  为类内判别权值,  $(1 - W_{ij})$  为类间判别权值;  $\mathbf{J}_k(\mathbf{x}_i)$  为  $\mathbf{x}_i$  的光谱域,  $k$  为光谱近邻参数;  $\mathbf{x}_j \in \mathbf{J}_k(\mathbf{x}_i)$  表示当构建光谱近邻图时, 像元  $\mathbf{x}_j$  在  $\mathbf{x}_i$  的光谱  $k$  邻域内,  $\mathbf{x}_j \notin \mathbf{J}_k(\mathbf{x}_i)$  表示像元  $\mathbf{x}_j$  不在  $\mathbf{x}_i$  的光谱  $k$  邻域内。该权重值示意图如图 1 所示, 其中不同的形状表示不同类别地物, 虚线圆圈表示以中间方形(像元)为圆心, 以光谱最大相似度为半径的光谱相似性近邻图; 圆内的像元表示与中心像元同谱, 即光谱相似; 圆外的像元表示与中心像元异谱, 即光谱不相似。

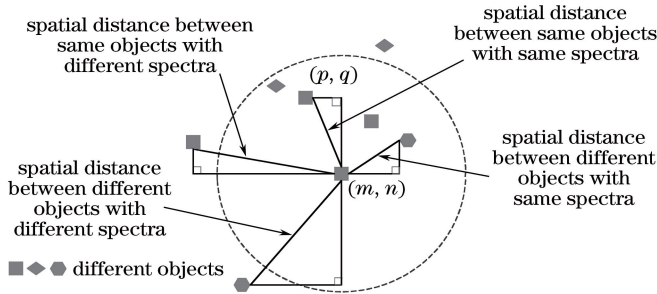


图 1 地物空间和像元近邻关系示意图

Fig. 1 Diagram of the neighborhood relationship between space and pixel of objects

基于空间信息和光谱信息构建的权重关系的优点: 1)  $0 \leq W_{ij} \leq 1, 1 \leq (1 + W_{ij}) \leq 2, 0 \leq (1 - W_{ij}) \leq 1$ , 类间判别权值和类内判别权值将像元冗余信息缩小到一定的范围之内; 2) 由(1)式可知, 随着欧氏距离的减小, 局部权重值增大; 类间判别权值减小, 类内判别权值增大; 使得同类像元的相似权值增大, 而非同类像元的相似权值相应减少。

该权重关系融合了光谱信息和空间信息, 提高了同类地物尤其是光谱相似的同类地物间的相似性及异类地物的差异性, 并更加全面地考虑了光谱不相似的地物间的空间关系及光谱差异性。

#### 3.2 考虑空间相关性的半监督 LPP 特征提取算法原理

本研究构建的权重关系, 基于光谱域信息, 同时利用像元的类标信息及像元间的相对空间位置信息, 给出考虑空间相关性的半监督局部保持的目标函数, 即

$$\min_{\mathbf{A}} \frac{1}{2} \sum_{i,j=1}^n (\mathbf{A}^T \mathbf{x}_i - \mathbf{A}^T \mathbf{x}_j) (\mathbf{A}^T \mathbf{x}_i - \mathbf{A}^T \mathbf{x}_j)^T q_{ij} =$$

$$\min_{\mathbf{A}} \frac{1}{2} \sum_{i=1}^n (\mathbf{A}^T \mathbf{x}_i D_{ii}^* \mathbf{x}_i^T \mathbf{A} - \mathbf{A}^T \mathbf{XQX}^T \mathbf{A}) =$$

$$\begin{aligned} \min_{\mathbf{A}} \frac{1}{2} \sum_{i=1}^n (\mathbf{A}^T \mathbf{X} \mathbf{D}^* \mathbf{X}^T \mathbf{A} - \mathbf{A}^T \mathbf{X} \mathbf{Q} \mathbf{X}^T \mathbf{A}) = \\ \min_{\mathbf{A}} \frac{1}{2} \sum_{i=1}^n \mathbf{A}^T \mathbf{X} (\mathbf{D}^* - \mathbf{Q}) \mathbf{X}^T \mathbf{A} = \\ \min_{\mathbf{A}} \frac{1}{2} \sum_{i=1}^n \mathbf{A}^T \mathbf{X} \mathbf{L}^* \mathbf{X}^T \mathbf{A}, \end{aligned} \quad (7)$$

式中:矩阵  $\mathbf{D}^*$  为  $\mathbf{Q}$  的对角矩阵;  $D_{ii}^* = \sum_j q_{ij}$ ;  $\mathbf{L}^* = \mathbf{D}^* - \mathbf{Q}$  为拉普拉斯矩阵。

引入约束条件  $\mathbf{Y} \mathbf{D}^* \mathbf{Y}^T = \mathbf{A}^T \mathbf{X} \mathbf{D}^* \mathbf{X}^T \mathbf{A} = \mathbf{I}$ , 则(7)式可以转化为 LPP-SCSSFE 求解最小值的数学模型

$$\arg \min_{\mathbf{A}} \mathbf{A}^T \mathbf{X} \mathbf{L}^* \mathbf{X}^T \mathbf{A}, \text{ when } \mathbf{A}^T \mathbf{X} \mathbf{D}^* \mathbf{X}^T \mathbf{A} = \mathbf{I}. \quad (8)$$

引入拉格朗日乘子值  $\lambda$  将(8)式转换为拉格朗日函数

$$\mathbf{L} = \mathbf{A}^T \mathbf{X} \mathbf{L}^* \mathbf{X}^T \mathbf{A} - \lambda (\mathbf{A}^T \mathbf{X} \mathbf{D}^* \mathbf{X}^T \mathbf{A} - \mathbf{I}). \quad (9)$$

根据(9)式求拉格朗日函数对变量  $\mathbf{A}$  的偏导数为

$$\partial \mathbf{L} / \partial \mathbf{A} = 2 \mathbf{X} \mathbf{L}^* \mathbf{X}^T \mathbf{A} - 2 \lambda \mathbf{X} \mathbf{D}^* \mathbf{X}^T \mathbf{A}. \quad (10)$$

令  $\partial \mathbf{L} / \partial \mathbf{A} = 0$ , 求解此时转换矩阵  $\mathbf{A}$  中的每个特征向量。广义特征值方程为

$$\mathbf{X} \mathbf{L}^* \mathbf{X}^T \mathbf{a}_i = \lambda_i \mathbf{X} \mathbf{D}^* \mathbf{X}^T \mathbf{a}_i. \quad (11)$$

利用转换矩阵  $\mathbf{A}$  对原数据集  $\mathbf{X}$  进行特征转换, 则有

$$\mathbf{x}_i \rightarrow \mathbf{y}_i = \mathbf{A}^T \mathbf{x}_i. \quad (12)$$

### 3.3 考虑空间相关性的半监督 LPP 特征提取算法步骤

本研究提出的考虑空间相关性的半监督 LPP 特征提取算法的步骤如下。

输入: 数据集  $\mathbf{X} = \{(\mathbf{x}_i, c_i)\}_{i=1}^l \cup \{\mathbf{x}_i\}_{i=l+1}^N$ , 参数为  $k$ ;

输出: 转换矩阵  $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \cdots \mathbf{a}_d]$  及低维子空间  $\mathbf{Y} = \mathbf{A}^T \mathbf{X} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N\}$ 。

算法的实现步骤如下。

步骤 1: 根据(1)式及像元的类标信息和空间关系构造新的同物异谱、同谱异物像元权重矩阵  $\mathbf{Q}$ ;

步骤 2: 计算权重矩阵的对角矩阵  $\mathbf{D}^*$  以及拉普拉斯矩阵  $\mathbf{L}^*$ ;

步骤 3: 根据(6)式构建的半监督权重关系构造考虑空间相关性的半监督局部保持的目标函数(7)式;

步骤 4: 根据(7)式确定 LPP-SCSSFE 的数学模型(8)式;

步骤 5: 求解广义特征方程(11)式;

步骤 6: 根据(11)式确定转换矩阵  $\mathbf{A} = [\mathbf{a}_1 \mathbf{a}_2 \cdots$

$\mathbf{a}_d]$  以及低维子空间  $\mathbf{Y} = \mathbf{A}^T \mathbf{X} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N\}$ 。

## 4 实验与分析

### 4.1 实验数据选择

为了检验 LPP-SCSSFE 算法的有效性, 采用 Indian Pines 数据集和 Pavia University 数据集进行实验, 简介如下:

1) Indian Pines 数据集是于 1992 年 6 月由 AVIRIS 传感器在美国印第安纳州的一块印度松树测试地获得的。图 2 为 Indian Pines 中的高光谱图像(145 pixel  $\times$  145 pixel)。该数据集共有 16 种不同类型地物, 10249 个像元, 220 个波段。由于本数据集有几类地物的像元数比较少, 因此, 本研究选择 10 类(像元数大于 300)地物进行实验。

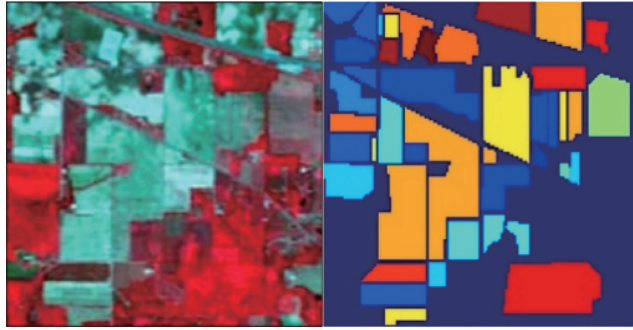
2) Pavia University 数据集是于 2002 年采用反射光学系统成像光谱仪获取得到的, 覆盖帕维亚大学的整个区域, 图 3 为 Pavia University 中的高光谱图像(610 pixel  $\times$  340 pixel)。该数据集共有 9 种不同类型地物, 每类地物在光谱覆盖范围为 0.43 ~ 0.86  $\mu\text{m}$  的区域共有 115 个波段, 排除各种噪声的干扰, 每类地物剩下 103 个波段。

### 4.2 实验分析

采用 LPP-SCSSFE 算法对 Indian Pines 和 Pavia University 中的高光谱图像进行特征提取, 并与无监督特征提取算法 PCA、LPP, 有监督特征提取算法 LSDA、NPE, 以及半监督特征提取算法 SSLPP 及 SSSDR-PCP 等进行对比。然后采用最近邻分类器(1-NN)<sup>[7,24]</sup> 和支持向量机(SVM)分类器对特征提取后的像元数据进行分类, 通过分类结果验证各特征提取算法的分类精度。实验中为了得到每种算法的最优分类效果, 利用交叉验证法调整各算法的参数, LPP 的光谱近邻参数  $k$  为 7; LSDA 的参数  $\mu$  为 0.1,  $k$  为 11; NPE 的  $k$  为 3; SSLPP 的  $k$  为 11; SSSDR-PCP 的  $k$  为 7,  $\alpha$  为 0.9; LPP-SCSSFE 的  $k$  为 7。最终采用了总体分类精度(OA)、平均分类精度(AA)和 Kappa 系数( $\kappa$ )<sup>[25]</sup> 作为评价指标。

#### 4.2.1 Indian Pines 数据实验

为了使特征提取算法得到的转换矩阵能涵盖所有地物的特征信息, 从 Indian Pines 数据集中每类地物的高光谱图像中随机选取 10, 20, 30 个像元点作为有类标样本, 然后从剩余像元中随机选取 3000 个无类标样本, 组成训练样本集, 其余的像元作为测试样本, 每种算法的实验重复 5 次, 最终取总体分类精度的平均值。



- alfalfa (46)      grass/pasture (483)      oats (20)      wheat (205)
- corn-notill (1428)      grass/trees (730)      soybean-notill (972)      woods (1265)
- corn-mintill (830)      grass/pasture-mowed (28)      soybean-mintill (2455)      bldg-grass-trees (386)
- corn (237)      hay-windrowed (478)      soybean-clean (593)      stone-steel (93)

图2 Indian Pines 中的高光谱图像及地面真值数据

Fig. 2 Hyperspectral image and ground truth data of Indian Pines

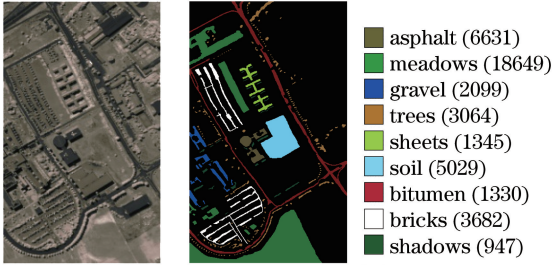


图3 Pavia University 中的高光谱图像及地面真值数据

Fig. 3 Hyperspectral image and ground truth data of Pavia University

图4给出了各算法在不同维数(0~50)下用1-NN分类器进行分类的平均总体分类精度,横坐标为投影转换后的维数,表1展示了不同特征提取算法在不同类标数、不同分类器(1-NN和SVM)情况下的最高总体分类精度及方差。表1中,用加粗着重突出在相同类标数及分类器的情况下分类精度的

最大值。

由图4和表1可以看出,对所有的特征提取方法,随着有标签样本数据集的增加,从样本数据中获得的先验知识也越全面,样本数据的总体分类精度也就越高。对于同样的标签数,LPP-SCSSFE、SSDR-PCP和SSLPP与LPP相比,虽然都保留了训练数据光谱近邻结构,但由于前3种方法利用了样本的标签信息,提取出了一些具有鉴别性的低维特征,所以分类精度都高于LPP。在1-NN分类器下,10,20,30个有标签训练样本的LPP-SCSSFE分类精度比SSDR-PCP算法分别提高了3.89%,3.51%和3.65%;在SVM分类器下,LPP-SCSSFE的分类精度比SSDR-PCP算法分别提高了4.95%,5.59%和5.04%。因此,本文算法不仅充分挖掘了图像数据的光谱特征和空间关系,而且大幅提高了分类精度。

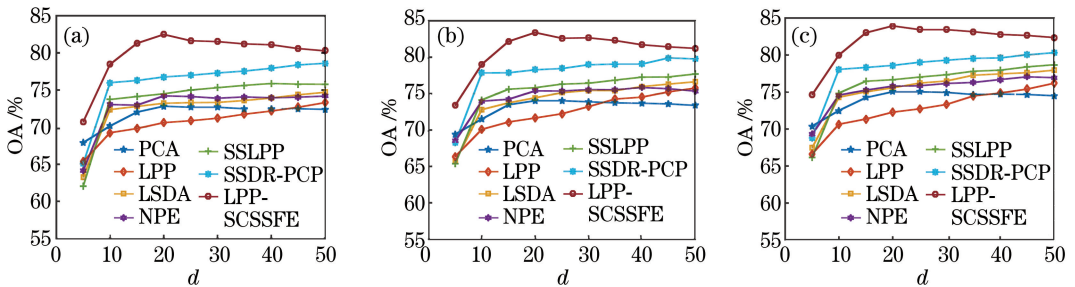


图4 不同算法对 Indian Pines 数据集中不同类标的总体分类精度。(a) 10 个类标样本;(b) 20 个类标样本;(c) 30 个类标样本

Fig. 4 Overall classification accuracy of different labels in Indian Pines dataset with different algorithms. (a) 10 labels; (b) 20 labels; (c) 30 labels

表 1 不同算法对 Indian Pines 数据集中不同类标的最高总体分类精度(平均值±方差/%)

Table 1 Highest overall classification accuracy of different labels in Indian Pines dataset with different algorithms (average±standard variance /%)

Algorithm	Highest overall classification accuracy					
	10 labels		20 labels		30 labels	
	1-NN	SVM	1-NN	SVM	1-NN	SVM
PCA	72.81±0.59	75.02±1.43	74.03±0.27	75.94±0.82	75.04±0.51	76.59±1.38
LPP	73.29±0.20	76.96±0.78	75.76±0.14	77.51±1.05	76.17±0.44	78.89±0.67
LSDA	74.69±2.67	78.20±3.69	76.61±1.92	80.92±1.90	77.96±2.84	81.15±2.17
NPE	74.21±2.20	73.21±2.69	75.84±1.25	75.16±1.27	77.06±2.16	78.45±1.75
SSLPP	75.89±1.54	77.25±1.54	77.70±1.47	81.38±1.59	78.68±1.95	82.09±0.70
SSDR-PCP	78.59±0.50	79.67±0.62	79.89±0.39	80.89±0.73	80.30±0.63	82.46±0.78
LPP-SCSSFE	<b>82.48±0.74</b>	<b>84.62±0.78</b>	<b>83.40±0.42</b>	<b>86.48±0.67</b>	<b>83.95±0.59</b>	<b>87.50±0.37</b>

表 2 列出了当有标签训练样本数为 30 个 (1-NN分类器)时,不同算法对 Indian Pines 数据集中各类地物的分类精度,表中用加粗表示各类地物分类精度的最大值。从表中可以看出:

表 2 不同算法对 Indian Pines 数据集中各类地物的分类精度

Table 2 Classification accuracies of different types of samples in Indian Pines dataset with different algorithms

Class	Classification accuracy /%						
	PCA	LPP	LSDA	NPE	SSLPP	SSDR-PCP	LPP-SCSSFE
1	68.17	63.88	64.84	60.93	75.39	<b>76.94</b>	73.46
2	63.00	61.38	65.63	66.25	62.88	62.25	<b>66.75</b>
3	85.07	83.40	64.46	80.57	84.26	85.35	<b>91.39</b>
4	89.29	88.29	90.86	88.86	93.00	93.86	<b>94.86</b>
5	97.11	96.33	95.42	97.75	<b>100.00</b>	98.88	<b>100.00</b>
6	66.91	67.09	68.97	68.20	72.82	<b>74.76</b>	70.48
7	74.72	74.19	77.07	77.37	77.74	79.57	<b>81.28</b>
8	52.66	53.11	67.00	62.27	56.09	55.12	<b>68.74</b>
9	90.31	90.96	94.37	95.95	90.65	91.01	<b>97.67</b>
10	52.58	51.12	65.96	60.28	62.51	61.12	<b>68.54</b>
OA	75.04	76.17	77.96	77.06	78.68	80.30	<b>83.95</b>
AA	73.98	72.98	75.46	75.84	77.53	77.89	<b>81.32</b>
$\kappa$	72.38	72.65	72.50	75.35	74.22	76.70	<b>80.99</b>

1) 与其他半监督算法对比,对比 LPP-SCSSFE 与 SSLPP、SSDR-PCP 可知,同一地物(除 1 和 6)的分类精度都相对提高。表 2 中第 8 类地物(Soybean-clean), LPP-SCSSFE 的分类精度比 SSLPP 提高了 12.65%,比 SSDR-PCP 提高了 13.62%。在该高光谱图像源数据中,第 3 类地物和第 10 类地物呈现同谱异物现象,通过本文 LPP-SCSSFE 算法处理后的分类精度高于 SSLPP、SSDR-PCP 算法处理后的精度。对比结果表明:LPP-SCSSFE 综合考虑样本之间的光谱相似性和空间相关性以及光谱不相似的地物间的差异性,极大提升了图像数据的空间区分度,可有效地获得更好的分类结果。

2) 与其他监督算法对比,对比 LPP-SCSSFE 与 LSDA、NPE 可知,LPP-SCSSFE 算法的 AA 和 Kappa 系数都优于 LSDA、NPE 这两种算法。这有

效地验证了本研究基于半监督的思想、利用光谱信息和空间信息构建的权重关系的可行性,发挥了算法中嵌入了无类标样本信息的能力,弥补了 LSDA 和 NPE 算法中在此方面的不足,使得提出的 LPP-SCSSFE 算法在标签训练样本较少的情况下,依然能够获得更高的分类效果。

3) 与其他无监督算法对比,对比 LPP-SCSSFE 与 PCA、LPP 可知,LPP-SCSSFE 算法的分类结果优于 PCA、LPP 这两种特征提取算法。它有效地利用了样本的类标信息,改进了样本间的相似性,显著地提升了样本的分类效果,有效地验证了半监督思想的优越性和有效性以及样本空间信息的重要性。

4) 由图 4 可知,当类标数为 30,维数为 20 时,所有地物的总体分类精度最高。当类标数为 30,由表 2 可知,本文 LPP-SCSSFE 算法对该数据集进行特征提取处理后,大部分地物的分类结果优于其他

的特征提取算法处理后的分类结果。由此也可以说明,图4中每类地物取一定数量的样本组成有类标数据集进行特征提取得到的转换矩阵是可行的。

#### 4.2.2 Pavia University 数据实验

为了使特征提取算法得到的转换矩阵能涵盖所有地物的特征信息,从 Pavia University 数据集中的每类地物的高光谱图像中随机选取 10, 20, 30 个像元作为有类标样本,然后从剩余像元中随机选取 3000 个无类标样本,组成训练样本集,其余的像元

作为测试样本,每种算法的实验重复 5 次,最终取总体分类精度的平均值。

图5给出了各种算法在不同维数(0~50)下用 1-NN 进行分类的平均总体分类精度,横坐标为投影转换后的维数,表3展示了不同特征提取算法在不同类标数、不同分类器(1-NN 和 SVM)情况下的最高总体分类精度及方差。表3中用加粗着重突出在相同类标数及分类器的情况下分类精度的最大值。

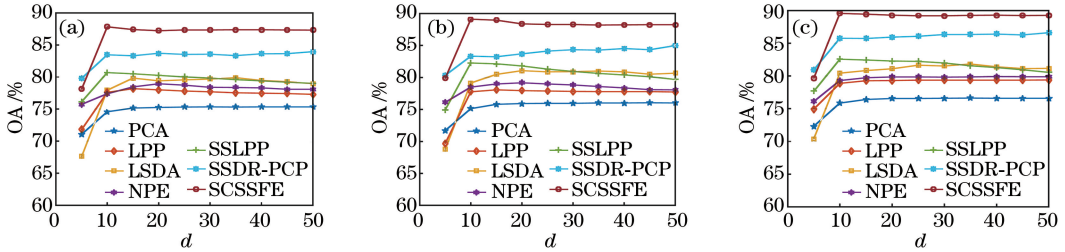


图5 不同算法对 Pavia University 数据集中不同类标的总体分类精度。(a) 10 个类标样本; (b) 20 个类标样本; (c) 30 个类标样本

Fig. 5 Overall classification accuracy of different labels in Pavia University dataset with different algorithms. (a) 10 labels; (b) 20 labels; (c) 30 labels

表3 不同算法对 Pavia University 数据集中不同类标的最高总体分类精度(平均值±方差/%)

Table 3 Highest overall classification accuracy of different labels in Pavia University dataset with different algorithms (average±standard variance /%)

Algorithm	Highest overall classification accuracy					
	10 labels		20 labels		30 labels	
	1-NN	SVM	1-NN	SVM	1-NN	SVM
PCA	75.36±0.09	75.83±1.49	76.03±0.43	75.32±1.55	76.62±0.17	76.37±2.99
LPP	78.19±1.08	78.22±0.48	78.36±0.31	78.51±1.17	79.40±0.99	79.47±0.48
LSDA	79.90±2.55	81.36±2.97	81.07±2.48	83.72±2.42	81.85±2.14	84.44±2.15
NPE	78.96±2.03	78.92±2.08	79.18±2.43	79.37±2.82	79.92±1.72	81.00±3.05
SSLPP	80.68±0.42	81.22±0.42	82.26±0.43	83.67±0.60	82.57±0.40	84.29±0.23
SSDR-PCP	83.96±0.39	83.18±0.84	84.99±0.31	84.23±0.96	86.61±0.12	85.38±1.28
LPP-SCSSFE	<b>87.84±0.72</b>	<b>88.11±0.47</b>	<b>89.10±0.65</b>	<b>90.04±0.12</b>	<b>89.57±0.40</b>	<b>91.29±0.23</b>

由图5和表3可以看出,对所有的特征提取方法,随着有标签样本数据集的增加,从样本数据中获得的先验知识也越全面,样本数据的总体分类精度也就越高。由于 PCA 和 LPP 为无监督特征提取算法,未利用标签信息,因此,整体的分类精度较低。而半监督特征提取算法 SSLPP、SSDR-PCP 和 LPP-SCSSFE 都保留了图像数据间的光谱邻域关系,且利用了少量的有标签样本数据,提高了总体分类精度,弥补了无监督特征提取的缺陷。在 SVM 分类器下,10, 20, 30 个有标签训练样本的 LPP-SCSSFE 分类精度比 SSSDR-PCP 的分别提高了 4.93%, 5.81% 和 5.91%。LPP-SCSSFE 充分利用了类内判别权值和类间判别权值,引入了像元的空

间信息,明确了光谱不相似的地物间的光谱距离,更有利于发现高光谱图像的内部结构,因此, LPP-SCSSFE 的总体分类精度有明显的提升。

表4列出了当有标签训练样本数为 30 个(1-NN)时,不同算法对 Pavia University 数据集中各类地物的分类精度,表中用加粗表示各类地物分类精度的最大值。从表中可以看出:

1) 与其他半监督算法对比,对比 LPP-SCSSFE 与 SSLPP、SSDR-PCP 可知,同一地物(除 4)的分类精度都相对提高。对第 3 类(Gravel), LPP-SCSSFE 的分类精度比 SSLPP 提高了 9.41%, 比 SSSDR-PCP 提高了 9.67%。对第 6 类(Bare Soil), LPP-SCSSFE 的分类精度比 SSSDR-PCP 提高了

表4 不同算法对 Pavia University 数据集中各类地物的分类精度

Table 4 Classification accuracies of different types of samples in Pavia University dataset with different algorithms

Class	Classification accuracy /%						
	PCA	LPP	LSDA	NPE	SSLPP	SSDR-PCP	LPP-SCSSFE
1	81.91	85.34	75.05	74.20	83.64	85.67	<b>86.47</b>
2	92.61	93.64	89.92	86.17	95.79	95.74	<b>96.65</b>
3	71.44	70.61	65.20	64.62	72.93	72.67	<b>82.34</b>
4	84.28	86.26	94.03	<b>94.13</b>	91.83	91.29	93.38
5	99.54	99.35	99.70	99.39	99.77	99.54	<b>99.85</b>
6	59.99	69.15	65.71	64.41	72.35	72.31	<b>78.44</b>
7	86.57	85.31	85.77	80.46	86.54	85.00	<b>90.75</b>
8	70.31	72.68	68.37	73.63	76.97	72.92	<b>78.77</b>
9	98.00	98.00	98.89	98.67	97.24	97.38	<b>98.91</b>
OA	76.62	79.40	81.85	79.92	82.57	86.61	<b>89.57</b>
AA	74.47	76.08	74.26	73.57	77.71	77.25	<b>80.56</b>
$\kappa$	73.73	76.63	78.48	76.61	78.96	83.01	<b>86.70</b>

6.13%。在该高光谱图像源数据中,第8类地物由于地理位置分布不同、环境因素的影响,有些像元可能被误分为背景,存在同物异谱的现象,经过本研究LPP-SCSSFE算法的处理,分类精度有相对的提升。对比结果表明:LPP-SCSSFE同时考虑像元光谱信息和空间信息,包括光谱相近的同类像元之间、光谱相近的不同类样本之间以及光谱不相近的同类样本之间的空间关系,更有利于提取样本数据的有用特征,可更有效地提高图像样本的分类结果。

2) 与其他监督算法对比,对比LPP-SCSSFE与LSDA、NPE可知,LPP-SCSSFE的AA和Kappa系数分别比LSDA提高了6.30%和8.22%,比NPE提高了6.99%和10.09%,有效地验证了本研究提出的考虑空间相关性的半监督LPP高光谱图像特征提取算法的可行性。在算法中同时加入少量有类标样本信息和大量无类标样本信息,而LSDA和NPE算法中的所有样本信息都加入了类标,因此,本文算法降低了LSDA和NPE算法中类标信息的成本,但依然能够获得较高的分类精度。

3) 与其他无监督算法对比,对比LPP-SCSSFE与PCA、LPP可知,对第3类(Gravel),LPP-SCSSFE的分类精度比PCA提高了10.90%,比LPP提高了11.73%。实验结果显示:本文算法的总体分类精度、平均分类精度及Kappa系数都优于这两种特征提取算法。由此分析,本文算法有效地引入类标信息,并利用样本的光谱相似性和空间相关性,及构造的同物异谱、同谱异物像元间的权重关系,明显提高了图像数据特征提取后的分类精度。

4) 由图5可知,当类标数为30,维数为10时,

所有地物的总体分类精度最高。当有类标数为30时,由表4可知,本文LPP-SCSSFE算法对该数据集进行特征提取处理后,除第4类地物,其他地物的分类精度均优于其他的特征提取算法处理后的精度。由此也可以说明,图5中每类地物取一定数量的样本组成有类标数据集进行特征提取得到的转换矩阵是合理的。

## 5 结 论

LPP算法能够保持图像数据的局部信息,但是没有利用图像样本的类标信息;而现有的基于LPP的半监督特征提取算法利用了光谱近邻信息和类标信息,但是没有充分考虑图像数据空间特征且忽视了光谱不相似的地物间的差异性。基于此问题,本研究提出一种考虑LPP的空间相关性的半监督特征提取(LPP-SCSSFE)算法。该算法的主要内容是将高维数据投影到低维空间,考虑数据的光谱信息和空间信息,减小低维空间中同类像元的距离,提高不同类像元的距离,继而提高样本数据的分类精度。通过在Indian Pines数据集及Pavia University数据集上随机选取10,20,30个有标签样本及3000个无标签样本作为训练样本,剩余样本作为测试集,最终对比不同特征提取算法的实验结果。在Indian Pines数据集上,LPP-SCSSFE的最高总体分类精度达到87.50%,Kappa系数为80.99%;在Pavia University数据集上,LPP-SCSSFE的最高总体分类精度达到91.29%,Kappa系数为86.70%,都高于其他的特征提取算法,均提高了高光谱图像的分类精度。为了得到更好的分类效果,下一步将在保持图像的局部特性的情况下,考虑图像的全局特性。



## 参 考 文 献

- [1] Liu B, Yu X C, Zhang P Q, *et al.* Supervised deep feature extraction for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(4): 1909-1921.
- [2] Wang J N. Hyperspectral image classification based on joint sparse representation and morphological feature extraction [J]. *Laser & Optoelectronics Progress*, 2016, 53(8): 082801.  
王佳宁. 基于联合稀疏表示与形态特征提取的高光谱图像分类 [J]. *激光与光电子学进展*, 2016, 53(8): 082801.
- [3] Liao J S, Wang L G. Hyperspectral image classification method based on fusion with two kinds of spatial information [J]. *Laser & Optoelectronics Progress*, 2017, 54(8): 081002.  
廖建尚, 王立国. 两类空间信息融合的高光谱图像分类方法 [J]. *激光与光电子学进展*, 2017, 54(8): 081002.
- [4] Huang H, He K, Zheng X L, *et al.* Spatial-spectral feature extraction of hyperspectral image based on deep learning [J]. *Laser & Optoelectronics Progress*, 2017, 54(10): 101001.  
黄鸿, 何凯, 郑新磊, 等. 基于深度学习的高光谱图像空-谱联合特征提取 [J]. *激光与光电子学进展*, 2017, 54(10): 101001.
- [5] Hou B H, Yao M L, Wang R, *et al.* Spatial-spectral semi-supervised local discriminant analysis for hyperspectral image classification [J]. *Acta Optica Sinica*, 2017, 37(7): 0728002.  
侯榜煊, 姚敏立, 王榕, 等. 面向高光谱图像分类的空谱半监督局部判别分析 [J]. *光学学报*, 2017, 37(7): 0728002.
- [6] Li D R, Zhang L P, Xia G S. Automatic analysis and mining of remote sensing big data [J]. *Acta Geodaetica et Cartographica Sinica*, 2014, 43(12): 1211-1216.  
李德仁, 张良培, 夏桂松. 遥感大数据自动分析与数据挖掘 [J]. *测绘学报*, 2014, 43(12): 1211-1216.
- [7] Lu G F, Wang Y, Zou J, *et al.* Matrix exponential based discriminant locality preserving projections for feature extraction [J]. *Neural Networks*, 2018, 97: 127-136.
- [8] Sellami A, Farah I R. A spatial hypergraph based semi-supervised band selection method for hyperspectral imagery semantic interpretation [J]. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 2016, 10(10): 1839-1846.
- [9] Zhang B. Advancement of hyperspectral image processing and information extraction [J]. *Journal of Remote Sensing*, 2016, 20(5): 1062-1090.  
张兵. 高光谱图像处理与信息提取前沿 [J]. *遥感学报*, 2016, 20(5): 1062-1090.
- [10] Gan Y H, Luo F L, Liu J H, *et al.* Feature extraction based multi-structure manifold embedding for hyperspectral remote sensing image classification [J]. *IEEE Access*, 2017, 5: 25069-25080.
- [11] Lodha S P, Kamapur S. Dimensionality reduction techniques for hyperspectral images [J]. *International Journal of Application or Innovation in Engineering & Management (IJAEM)*, 2014, 3(10): 92-99.
- [12] Zhang X R, Huyan N, Zhou N, *et al.* Semi-supervised sparse dimensionality reduction for hyperspectral image classification [C] // 2016 IEEE Region 10 Conference, November 22-25, 2016, Singapore. New York: IEEE, 2016: 2830-2833.
- [13] Xue S Z, Tan R, Chen X H. Method of kernel-based semi-supervised locality preserving projection [J]. *Journal of Computer Applications*, 2012, 32(8): 2235-2237, 2244.  
薛寺中, 谈锐, 陈秀宏. 基于核的半监督的局部保留投影降维方法 [J]. *计算机应用*, 2012, 32(8): 2235-2237, 2244.
- [14] Wold S, Esbensen K, Geladi P. Principal component analysis [J]. *Chemometrics and Intelligent Laboratory Systems*, 1987, 2(1/2/3): 37-52.
- [15] He X F. Locality preserving projections [D]. Chicago: University of Chicago, 2005.
- [16] Zhang J B, Wang J K, Cai X. Sparse locality preserving discriminative projections for face recognition [J]. *Neurocomputing*, 2017, 260: 321-330.
- [17] Wan M H, Lai Z H. Multi-manifold locality graph embedding based on the maximum margin criterion (MLGE/MMC) for face recognition [J]. *IEEE Access*, 2017, 5: 9823-9830.
- [18] Huang Y, Tang K, Sun Z. Locality sensitive discriminant analysis for classification of hyperspectral data [C] // 2014 7th International Congress on Image and Signal Processing, October 14-16, 2014, Dalian, China. New York: IEEE, 2014: 292-296.
- [19] He X F, Cai D, Yan S C, *et al.* Neighborhood

- preserving embedding[C]//Tenth IEEE International Conference on Computer Vision (ICCV'05), October 17-21, 2005, Beijing, China. New York: IEEE, 2005, 2: 1208-1213.
- [20] Wu Z Y, Hou B Y, Wang H B, *et al.* Semi-supervised sparsity neighboring preserving projection [J]. *Systems Engineering and Electronics*, 2018, 40(4): 934-940.  
吴振宇, 侯冰洋, 王辉兵, 等. 半监督稀疏近邻保持投影[J]. *系统工程与电子技术*, 2018, 40(4): 934-940.
- [21] Lu X H, He Z Y, Yi S Y, *et al.* Joint of locality-and globality-preserving projections [J]. *Signal, Image and Video Processing*, 2018, 12(3): 565-572.
- [22] Pan Y S, Wang P F, Huang H, *et al.* Hyperspectral remote sensing image classification based on SSLPP [J]. *Computer Science*, 2013(S2): 333-336, 373.  
潘银松, 王攀峰, 黄鸿, 等. 基于 SSLPP 算法对高光  
谱遥感影像分类[J]. *计算机科学*, 2013(S2): 333-336, 373.
- [23] Du W B, Lü M, Hou Q L, *et al.* Semisupervised dimension reduction based on pairwise constraint propagation for hyperspectral images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2016, 13(12): 1880-1884.
- [24] Huang H, Huang Y B. Improved discriminant sparsity neighborhood preserving embedding for hyperspectral image classification [J]. *Neurocomputing*, 2014, 136: 224-234.
- [25] He F, Wang R, Yu Q, *et al.* Feature extraction of hyperspectral images of weighted spatial and spectral locality preserving projection (WSSLPP) [J]. *Optics and Precision Engineering*, 2017, 25(1): 263-273.  
何芳, 王榕, 于强, 等. 加权空谱局部保持投影的高光谱图像特征提取[J]. *光学精密工程*, 2017, 25(1): 263-273.