

联合局部二值模式与K-最近邻算法的高光谱图像分类方法

赵晋陵^{1,2*}, 胡磊², 严豪², 储国民², 方艳², 黄林生^{1,2*}

(1. 安徽大学 农业生态大数据分析与应用技术国家地方联合工程研究中心, 安徽 合肥 230601;
2. 安徽大学 电子信息工程学院, 安徽 合肥 230601)

摘要:如何利用较少训练样本达到高分类精度已成为高光谱遥感领域的重要研究方向和极具挑战性任务。针对高光谱图像包含的丰富光谱与空间信息,提出了一种联合局部二值模式LBP(Local Binary Patterns)与K-最近邻KNN(K-Nearest Neighbors)算法的高光谱图像分类方法。首先,通过主成分分析PCA(Principle Component Analysis)对高光谱数据进行降维;然后,使用LBP提取降维后的高光谱图像空间纹理信息,将光谱与空间特征变量堆叠成一谱特征向量;最后,输入最近邻分类器得到分类结果。选取Pavia University、Indian Pines和Salinas三种公开高光谱数据集的训练集和测试集作为建模和验证数据源,选取KNN、基于径向基核函数的支持向量机(RBF-SVM)、核联合正交匹配追踪(Kernel Simultaneous Orthogonal Matching Pursuit, KSOMP)三种经典分类算法作为比较。在Pavia University与Indian Pines数据集中随机选取10%作为训练样本,总体精度和Kappa系数分别达到99.15%、98.87%和97.88%、97.58%;在Salinas数据集中随机选取2%作为训练样本,总体精度与Kappa系数为98.46%和98.29%。实验结果表明,在训练样本仅为数据集10%甚至2%的条件下,本文提出的方法仍可达到98%以上的分类精度,可满足训练样本难以获取的应用场景对高分类精度要求。

关键词:高光谱遥感;局部二值模式;K-最近邻;光谱特征;主成分分析
中图分类号:P237 **文献标识码:**A

Hyperspectral image classification combining local binary patterns and k-nearest neighbors algorithm

ZHAO Jin-Ling^{1,2*}, HU Lei², YAN Hao², CHU Guo-Min², FANG Yan², HUANG Lin-Sheng^{1,2*}

(1. National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University, Hefei 230601, China;
2. School of Electronics and Information Engineering, Anhui University, Hefei 230601, China)

Abstract: It is a highly important and challenging task to finish the high-accuracy hyperspectral image classification using fewer training samples. A novel hyperspectral image-based classification method (hereafter referred to as the LBP-SSKNN) was proposed by combining Local Binary Patterns (LBP) and K-Nearest Neighbors (KNN). First, the Principal Component Analysis (PCA) was used to reduce the dimension of hyperspectral image. Subsequently, the LBP was used to extract the spatial texture information and the spatial and spectral features were uniformly scaled to form the spatial-spectral vectors. Finally, the vectors were input into the KNN classifier to obtain the classification result. The training and test datasets of three popular open hyperspectral datasets were used to validate the proposed method, including Pavia University, Indian Pines and Salinas. the classification method was verified on three groups of hyperspectral remote sensing image datasets. In addition, three classic classifiers were also selected to compare the LBP-SSKNN, including Radial Basis Function Support Vector Machine (RBF-SVM) and Kernel Simultaneous Orthogonal Matching

收稿日期:2020-06-29,修回日期:2021-03-08

Received date:2020-06-29, Revised date:2021-03-08

基金项目:国家自然科学基金“基于无人机遥感影像融合的地块尺度小麦白粉病解析方法研究”(31971789);安徽省自然科学基金“小麦白粉病侵染风险评估体系构建与预测模型研究”(2008085MF184);安徽省高等学校自然科学研究项目(KJ2019A0030)

Foundation items: Supported by National Natural Science Foundation of China (31971789); Natural Science Foundation of Anhui Province (2008085MF184); Natural Science Research Project of Anhui Provincial Education Department (KJ2019A0030)

作者简介(Biography):赵晋陵(1981-),山西晋城人,男,副教授,主要从事遥感影像处理与作物病虫害遥感测报研究. E-mail: zhaojl@ahu.edu.cn

*通讯作者(Corresponding author): E-mail: zhaojl@ahu.edu.cn; linsheng0808@163.com

Pursuit (KSOMP). In the Pavia University dataset and Indian Pines dataset, the 10% of training dataset were randomly selected as the training samples. The overall accuracy (OA) and Kappa coefficient reach 99.15%, 98.87% and 97.88%, 97.58%, respectively. In the Salinas dataset, only the 2% of training dataset were randomly selected as the training samples, and the OA and Kappa coefficient reach 98.46% and 98.29%. The experimental results show that the OA of LBP-SSKNN method can still reach more than 98% under the 10% and even 2% of the training dataset. Our proposed method can satisfy the high-accuracy requirement due to limited training samples in practical application scenes.

Key words: hyperspectral remote sensing, local binary patterns (LBP), k-nearest neighbors (KNN), spatial and textural features, principle component analysis

引言

近年来,高光谱图像分类一直是一个非常活跃的研究领域和热点方向,已在环境制图,作物分析,植物和矿物勘探、生物和化学检测等多应用场景中,发挥了重要甚至是不可替代的作用^[1]。因此,如何利用高光谱数据的高分辨率“空一谱”信息,不断提高分类精度,成为全球科研人员不断追求的目标。相比于常用的多光谱图像,高光谱技术可以获得光谱间隔为 nm 级的上百个窄带光谱通道,获取更丰富的光谱信息,便于对研究对象进行精细探测和精确解析。同时,高光谱影像存在的多波段、高相关性、冗余性等严重影响了分类精度和运行速度,迫切需要进行光谱降维^[2]和波段选择^[3]。传统降维方法主要有无监督、有监督两类。例如:主成分分析(Principal Component Analysis, PCA)^[4]是一种无监督降维方法,在去除高光谱数据大量冗余信息的同时,能保留方差贡献率较大主成分的光谱信息。线性判别分析(Linear Discriminant Analysis, LDA)^[5]是监督方法中一种常用的降维方法,数据集的每个样本都有类别输出。通过将高维的高光谱数据采用降维映射到低维子空间,再将保留的光谱信息输入到分类器中实现分类。常用的分类方法有:最近邻(K-Nearest Neighbors, KNN)分类法^[6]、支持向量机(Support Vector Machine, SVM)^[7]、光谱角制图法(Spectral Angle Mapping, SAM)^[8]、极限学习机(Extreme Learning Machine, ELM)^[9]等。大量研究表明,相比于其它常用分类算法, KNN 更加简单、高效且易于实现。

一方面,由于目标空间分布的复杂性和光谱的异质性,仅用光谱特征进行分类往往无法达到较高的精度^[10]。另一方面,在高光谱遥感图像中,相邻像素可能具有相同的上下文属性。因此,在高光谱图像分类中集成光谱和空间信息是提高分类精度的有效方法^[11]。例如,马尔科夫随机场(Markov Random Field, MRF)^[12]、灰度共生矩阵(Gray-Level

Co-occurrence Matrix, GLCM)^[13]、小波变换^[14]、Gabor 变换^[15]等。MRF 主要对图像像素的空间信息进行建模, Tarabalka 等^[16]和 Zhang 等^[17]使用 MRF 合并空间信息,建立结合 SVM 的 SVM-MRF 算法细化分类结果。黄鸿等^[18]将加权空谱距离(Weighted Spatial-Spectral Distance, WSSD)的相似性度量方法与 KNN 分类器结合(WSSD-KNN)有效提高了高光谱图像的地物分类精度。KSOMP^[19]是在 SOMP^[20]算法上的改进,充分利用了光谱和空间信息,采用稀疏表示高光谱的分类方法。由于高光谱数据在原始特征空间线性不可分,可以通过核方法将其映射到线性可分的高维空间,以便提高分类准确度。上述文献表明,考虑空间信息的高光谱图像分类算法优于传统单特征分类方法。

因此,高光谱图像分类精度的提升既要通过降维去除冗余信息,提高训练速度,也要最大限度融合光谱信息与空间信息。经对比研究发现,局部二值模式(Local Binary Patterns, LBP)是一种非参数方法,能有效提取高光谱数据中的纹理信息,具有旋转不变性和灰度不变性的优点^[21]。针对低复杂度分类算法和高分类精度的实际需求,本文提出了一种结合 LBP 与 KNN 分类器的高光谱空谱联合分类算法(LBP-SSKNN)。基于 Pavia University、Indian Pines 和 Salinas 三种主流公开高光谱数据集,比较 KNN、支持向量机(Support Vector Machine, SVM)和核联合正交匹配追踪(Kernel Simultaneous Orthogonal Matching Pursuit, KSOMP)三种经典分类算法,验证 LBP-SSKNN 算法的分类精度。

1 算法原理

1.1 光谱与空间特征提取

特征提取包括光谱特征和空间特征两个并行的提取步骤。将每个像素的光谱和空间特征叠加成一维向量。对于光谱信息的提取,采用 PCA 降维法保留包含绝大部分方差的维度特征,忽略方差贡

献率较小的维度特征,代表高光谱原始数据的光谱信息,从而减少计算成本^[22]。令输入的高光谱数据集为:

$$X = \{x_1, x_2, \dots, x_n\}, \quad (1)$$

式(1)中, $x_i = (x_{i1}, x_{i2}, \dots, x_{iB})^T (i = 1, 2, \dots, n)$, 代表每个像素样本的光谱矢量, n 为像素量, B 为波段数。

将 X 的每一行(即每一波段的像素样本)进行零均值化,即减去这一行的均值。另 $X' = x_i - a$, a 为每行均值;再对 X' 的协方差矩阵 $C_{B \times B}$ 采用特征值分解法求其特征值与特征向量。即:

$$C_{B \times B} = U \Lambda U^T, \quad (2)$$

式(2)中, U 为特征向量矩阵, $\Lambda = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_B]$ 为协方差矩阵 $C_{N \times N}$ 的特征值对角矩阵,且 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_B$, 都为非负整数。将前 p 个最大特征值对应的特征向量从上到下按行排列成矩阵 $H_{K \times B}$, 则高光谱降维后的数据 Y 为:

$$Y = (HX)^T, \quad (3)$$

式(3)中, $Y = (y_1, y_2, \dots, y_p)$, y_i 为第 i 个主成分的列向量。PCA 降维能去除高维度数据中不重要的特征,保留主要特征,随着主成分 p 个数的增加, y_i 所含信息量越来越少,虽然高光谱数据总体信息量也会增加,但同样可能会带来大量冗余信息。在最大限度保留信息量的同时,减少计算成本,提高分类精度。将在下文实验部分通过总体精度与 Kappa 系数确定最优主成分 p 的数值。

LBP 是由 Ojala 等^[21]提出的一种描述图像局部纹理的特征算子。它除了对图像强度水平变化具有简单性和鲁棒性之外,计算效率在纹理特征提取方面也具有明显优势。本文采用该算法提取高光谱数据的空间纹理特征。经典 LBP 采用 3×3 方形窗口的 8 邻域描述图像纹理特征,但难以满足不同尺度和频率要求,因此 Ojala 等人对其进行了改进,将 3×3 的方形窗口扩展到任意圆形领域。设输入图像中的中心像素 (x_c, y_c) 具有灰度值 g_c , 在其半径为 r 的圆周路径上,有 s 个均匀分布的采样点,每个采样点灰度值记为 $g_i (i = 1, \dots, p)$, 如图 1(a)、(b)所示。通过比较中心像素点 (x_c, y_c) 的灰度值与在其周围邻域距离为 r 处的 s 个采样像素点的灰度值,沿着圆周路径顺时针或逆时针旋转,如果中心像素值 g_c 具有比邻近像素采样点更高的灰度值,则为邻近像素分配 0 值,否则为 1,即生成 s 位二进制数,如图 1(c)所示。

此时,该窗口沿着顺时针组合成一个八位的二

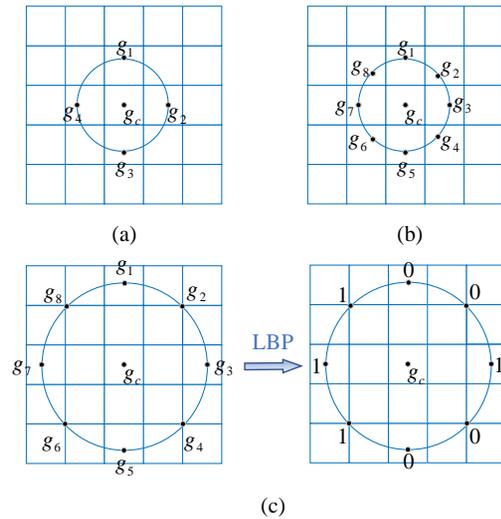


图1 (a) $LBP_{s=4, r=1}$, (b) $LBP_{s=8, r=1}$, (c) LBP 计算示意图

Fig. 1 (a) $LBP_{s=4, r=1}$, (b) $LBP_{s=8, r=1}$, (c) schematic diagram of LBP

进制数,然后再转换成十进制数: $(00100111)_2 = 39$, 即该 LBP 码值可以反映该窗口的区域纹理特征信息。对于 $s=8$ 时, LBP 可以生成 256 个不同的二进制模式, LBP 可以定义为:

$$LBP_{s,r}(x_c, y_c) = \sum_{i=1}^{s-1} f(g_i - g_c) 2^i, \quad (4-1)$$

其中:

$$f(x) = \begin{cases} 0, & f < 0 \\ 1, & f \geq 0 \end{cases}, \quad (4-2)$$

式(4)中, (x_c, y_c) 为中心像素坐标, $f(g_i - g_c)$ 为符号函数, g_c 为中心像素灰度值, g_i 为邻域采样点灰度值。

具体实现流程为:首先,将输入的高光谱图像划分为 $w \times w$ 个局部子区域,其次对每个子区域内的中心像素点都提取 LBP 特征;然后,在每个子区域建立 LBP 特征的统计直方图,其横坐标表示 LBP 的类别,纵坐标表示每一类别出现的特征个数,每个子区域就可以用一个统计直方图进行描述;最后,再将每个子区域的 LBP 直方图级连起来,形成空间特征向量。值得注意的是,局部子区域 w 是用户自定义的参数,不同的 w 会有不同分类效果。如果 w 太大,会增加计算复杂度,否则提取的特征不能代表中心像素的空间特征。本文的 w 值将在实验部分进行测试和确定。

1.2 KNN 分类器

KNN 算法是一种基于欧几里得距离推断事物类别的分类方法^[23]。通过计算对象间距离作为各

个对象之间的非相似性指标,避免了对象之间的匹配问题,距离公式如下

$$D(x,y) = \sqrt{\sum_{k=1}^n (x_k - y_k)^2} \quad (5)$$

主要计算步骤为:首先,输入随机产生的训练集 (x_i, y_i) ,计算测试数据与各个训练数据之间的距离 $D(x, y)$;然后,按照距离的递增关系进行排序,选取距离最小的前 k 个点;最后,对近邻中类别结果进行投票,返回前 k 个点中投票率最高的类别作为测试数据的预测分类。

1.3 联合局部二值模式与K-最近邻算法的高光谱图像分类方法

光谱特征是区分不同地物类型的重要信息,而空间特征降低了类内方差,可以提高分类器的性能。因此,光谱和空间特征的结合能提供更全面和可靠的分类结果。本文提出了结合LBP与KNN的高光谱空谱联合分类方法(LBP-SSKNN),具体实现流程如图2所示。

详细分类步骤如下:

(1) 对原始高光谱影像进行归一化,再采用PCA对其降维,保留前 p 个方差贡献率最大的主成分,作为该高光谱数据的光谱特征,生成光谱特征向量;

(2) 将降维后的高光谱影像划分为 $w \times w$ 个局部子区域,对每个子区域使用LBP算法提取空间纹理特征,再将所有子区域提取的空间特征联接起来,生成空间特征向量,作为高光谱数据的空间特征;

(3) 将光谱与空间特征向量统一缩放到 $[-1, 1]$ 范围内,使缩放后特征矩阵的各个维度都具有相同重要性,加快训练速度;

(4) 对缩放后的光谱与空间特征向量进行矢量

叠加,具体来说,即将空间特征向量添加到光谱特征向量的末端,最终生成一个新的空-谱特征矩阵;

(5) 将生成的空-谱特征矩阵输入KNN分类器进行图像分类,对分类结果进行精度评价。

1.4 分类精度评价指标

本文采用了基于混淆矩阵计算的总体精度(Overall Accuracy, OA)、平均精度(Average Accuracy, AA)和Kappa系数作为评价指标^[24-25]。总体精度(OA)是模型在所有测试集上预测正确的样本个数与总样本数之间的比值。平均精度(AA)是模型对每类的分类精度与数据集总类别数的比值。OA与AA的计算公式如下:

$$OA = \frac{\sum_{i=1}^m x_{ii}}{\sum_{i=1}^m \sum_{j=1}^m x_{ij}} \quad (6)$$

$$AA = \frac{\sum_{i=1}^m \left(x_{ii} / \sum_{j=1}^m x_{ij} \right)}{m} \quad (7)$$

式(6)、(7)中, m 为类别数; x_{ii} 为第 i 类像元的正确分类数量,即混淆矩阵主对角线上的值; x_{ij} 为实际正确分类或错分的数量。

Kappa系数一般用于一致性检验,也可以用于衡量分类精度,其计算结果范围为 $-1 \sim 1$,但通常落在 $0 \sim 1$ 间。通常情况下,可划分为五种一致性等级:0.0~0.20极低的一致性(slight)、0.21~0.40一般的一致性(fair)、0.41~0.60中等的一致性(moderate)、0.61~0.80高度的一致性(substantial)和0.81~1.00几乎完全一致(almost perfect)。Kappa系数计算公式如下:

Kappa =

$$\frac{\left[N \sum_{i=1}^m x_{ii} - \sum_{i=1}^m (x_{i+} \times x_{+i}) \right]}{\left(N^2 - \sum_{i=1}^m (x_{i+} \times x_{+i}) \right)} \quad (8)$$

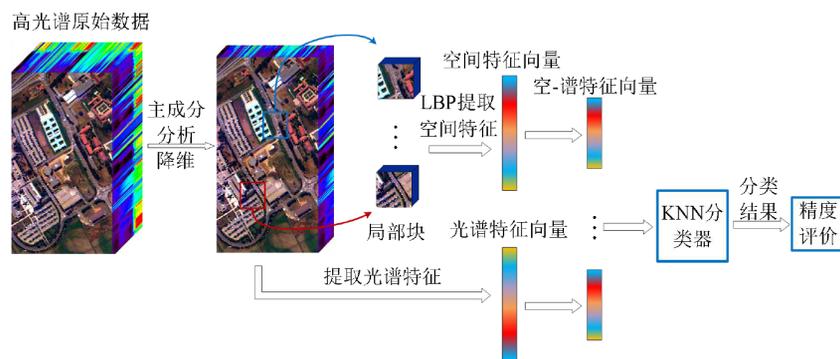


图2 LBP-SSKNN高光谱图像分类流程图

Fig. 2 Flowchart of hyperspectral image classification based on the LBP-SSKNN

式(8)中, m 为分类类别数; N 为用于精度评价的像元总数, x_{i+} 表示第 i 行的像元总数; x_{+i} 表示第 i 列的像元总和。

2 结果与分析

选取 Pavia University、Indian Pines 和 Salinas 三个典型公开高光谱数据集(www.ehu.eus/ccwintco/index.php?title=Hyperspectral_Remote_Sensing_Scenes)验证 LBP-SSKNN 方法的有效性和精度。计算机硬件环境为 Intel Corei5-7300HQ, CPU 2.50GHz, RAM 8GB, 软件环境为 Windows10(64)位操作系统, MATLAB 2016a。

2.1 实验数据

Pavia University 数据集。由德国的反射光学光谱仪 ROSIS-03 (Reflective Optics Spectrographic Imaging System) 在 2003 年获取的意大利帕维亚大学校园高光谱影像(图 3a, 由 R=60, G=30, B=2 假彩色合成)。图像的空间分辨率为 1.3 m, 大小为 610 像素×340 像素, 共有 42776 个样本, 但包含大量的背景像素, 地物种类有九种(图 3b 和表 1)。波长范围为 0.43~0.86 μm , 共有 115 个波段数, 其中 12 个波段由于受噪声影响被剔除, 本文使用剩下的 103 个光谱波段。

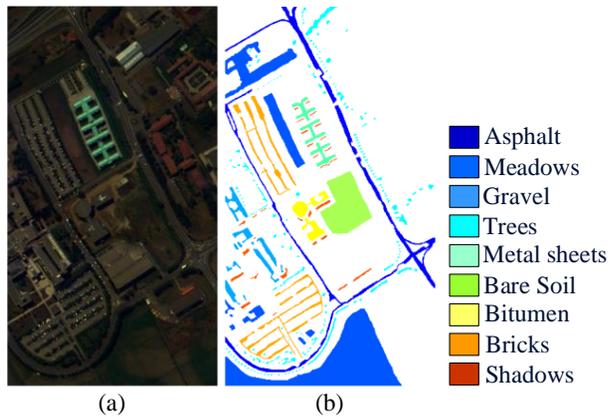


图3 Pavia University(a)假彩色影像(b)真实地物分布
Fig. 3 (a) False-color composite image (b) ground-truth classes for the Pavia University scene

Indian Pines 数据集。由可见光/红外成像光谱仪(AVIRIS)于 1992 年对美国印第安纳州西北部的一块印度松树进行拍摄(图 4a, 由 R=50, G=27, B=17 假彩色合成), 空间分辨率为 20 m, 容易产生影响分类精度的混合像元^[26]。图像尺寸为 145 像素×145 像素, 共 10249 个样本, 包含十六种地物类型(图 4b 和

表1 Pavia University 数据集样本信息

Table 1 Sample information of the Pavia University dataset

序号	地物类别	样本量
1	Asphalt	6 631
2	Meadows	18 649
3	Gravel	2 099
4	Trees	3 064
5	Painted metal sheets	1 345
6	Bare Soil	5 029
7	Bitumen	1 330
8	Self-Blocking Bricks	3 682
9	Shadows	947

表 2)。波长范围为 0.4~2.5 μm , 共有 220 个波段数, 剔除其中受噪声影响的 20 个波段, 本文使用剩下的 200 个光谱波段。

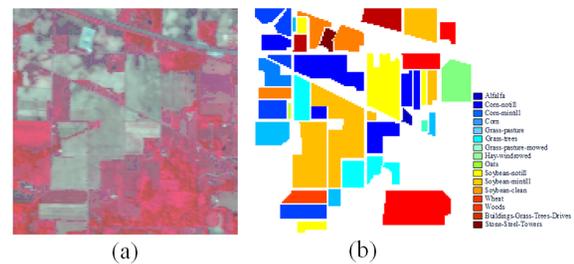


图4 Indian Pines(a)假彩色影像(b)真实地物分布

Fig. 4 (a) False-color composite image (b) ground truth classes for the Indian Pines scene

表2 Indian Pines 数据集样本信息

Table 2 Sample information for the Indian Pines dataset

序号	地物类别	样本量
1	Alfalfa	46
2	Corn-notill	1 428
3	Corn-mintill	830
4	Corn	237
5	Grass-pasture	483
6	Grass-trees	730
7	Grass-pasture-mowed	28
8	Hay-windrowed	478
9	Oats	20
10	Soybean-notill	972
11	Soybean-mintill	2 455
12	Soybean-clean	593
13	Wheat	205
14	Woods	1 265
15	Buildings-Grass-Trees-Drives	386
16	Stone-Steel-Towers	93

Salinas数据集。也是由AVIRIS拍摄。采集地点为美国加利福尼亚的Salinas山谷(图5a,由R=50, G=30, B=20假彩色合成), Salinas的空间分辨率为3.7 m, 图像尺寸为512像素×217像素, 共包含53829个样本, 含有十六种地物类型(图5b和表3)。原始数据包含224个波段, 剔除受噪声影响的20个波段, 使用剩下的204个波段。

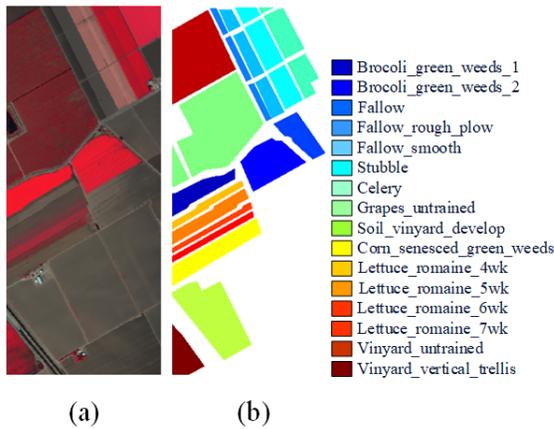


图5 Salinas假彩色影像(a), 与真实地物分布(b)

Fig. 5 False-color composite image (a) and ground truth classes, (b) for the Salinas scene

表3 Salinas数据集样本信息

Table 3 Sample information for the Salinas dataset

序号	地物类别	样本量
1	Broccoli_green_weeds_1	2 009
2	Broccoli_green_weeds_2	3 426
3	Fallow	1 976
4	Fallow_rough_plow	1 394
5	Fallow_smooth	2 678
6	Stubble	3 959
7	Celery	3 579
8	Grapes_untrained	11 271
9	Soil_vinyard_develop	6 203
10	Corn_senesced_green_weeds	3 278
11	Lettuce_romaine_4wk	1 068
12	Lettuce_romaine_5wk	1 927
13	Lettuce_romaine_6wk	916
14	Lettuce_romaine_7wk	1 070
15	Vinyard_untrained	7 268
16	Vinyard_vertical_trellis	1 807

2.2 参数设置

LBP-SSKNN方法的主要参数包括主成分个数 p , LBP算法的邻域半径 r 、近邻采样点个数 s , 划分的局部子区域大小 w , 分类器近邻数 k 。为提高实验结果的精确度和可靠性, 将每种算法重复10次, 每次

从每类地物样本随机选取10%训练样本, 其余作为测试样本, 将10次实验的分类精度取平均值作为结果。首先, 通过设置1~8个主成分数, 通过主成分贡献率与分类总体精度分析 p 对三个数据集分类结果的影响(表4)。

表4 三个数据集主成分贡献率对比

Table 4 Comparison of principal component contribution for the three datasets

主成分个数	Pavia University		Indian Pines		Salinas	
	贡献率/%	累计贡献率/%	贡献率/%	累计贡献率/%	贡献率/%	累计贡献率/%
1	58.32	58.32	68.49	68.49	74.47	74.47
2	36.10	94.42	23.53	92.02	23.53	98.00
3	4.43	98.86	1.49	93.52	1.13	99.14
4	0.30	99.16	0.82	94.34	0.54	99.68
5	0.21	99.37	0.69	95.03	0.17	99.85
6	0.18	99.54	0.52	95.55	0.06	99.91
7	0.12	99.67	0.40	95.95	0.02	99.93
8	0.07	99.74	0.36	96.31	0.01	99.95

如图6所示, 对于Salinas数据集, 主成分从1缓慢增加到2后, 随着 p 的不断增大, OA呈现稳定的趋势。对于Pavia University数据集, 当主成分个数从1增加到3时, OA随之增加, 此时前3个主成分累计贡献率接近99%(表4); 当主成分个数大于3时, OA有缓慢下降。对于Indian Pines数据集, 当 $p=4$ 时, 累计贡献率超过94%, OA达到最大值; 当 $p>4$ 时, OA也有一个下降过程, 并没有持续升高, 是因为 p 增加的同时, 虽然会增大光谱特征信息量, 但到一定程度时也会引入冗余特征。因此, 在允许损失信息量范围内, 选择一个相对较小的个数, 在降低计算复杂度同时, 也能保持较理想分类精度。

经过PCA降维后, 需要划分检测窗口以便提取LBP空间特征, 因此划分区域参数 w 也会对分类结果产生影响。以Indian Pines数据集为例, 设置 w 为3、5、7、9、11、13、15, 观察OA、AA、Kappa三个指标, 分析 w 对分类精度的影响(表5), 括号内为该组数据的标准偏差(以下简称标准差), 标准差数值越小表示分类精度越稳定, 反之亦然。分析表5中的数据可知, 随着 w 的增加, 分类精度缓慢上升, 当 $w=9$ 时, 达到最优的OA与Kappa, 分类效果最好, 且此时标准差最小; 当 $w>9$ 时, OA并没有持续升高。若 w 值过大, 则需要考虑不同类别的像素数, 可能会对分类精度产生负面影响。如果 w 过小, 提取的特征不能代表中心像素的空间纹理特征。因此, 对于In-

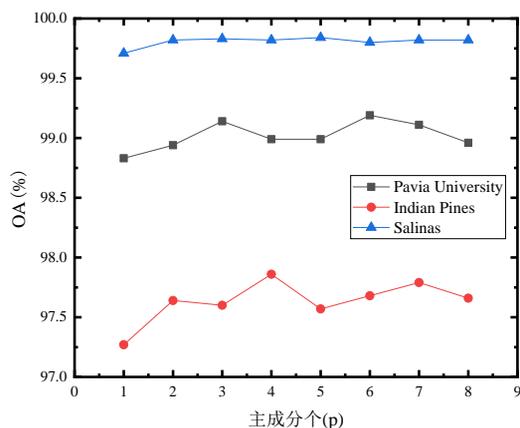


图6 主成分个数对分类精度的影响

Fig. 6 Influence on classification accuracies for the number of principal components

dian Pines 数据集的 LBP 特征提取设置 w 为 9 像素 \times 9 像素。由图 7 可知,在相同邻域采样点个数 s 条件下,不同邻域半径 r 对分类结果的影响较小;当 $(r, s)=(1, 8)$ 时,OA 达到最优值。随着 s 的增加,OA 并没有持续上升,总体变化幅度不大,但计算成本不断增加。因此,在对运算效率要求较高时,可适当减小 s ,降低计算复杂度。

近邻数 k 是影响 KNN 分类能力的主要因素,故

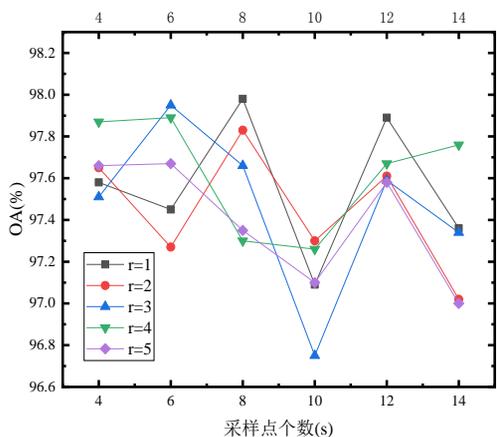


图7 Indian Pines 数据集参数 r 和 s 对分类精度的影响

Fig. 7 Influence on classification accuracies of the Indian Pines dataset for r and s

表5 参数 w 对 Indian Pines 数据集分类精度的影响分析

Table 5 Influence on classification accuracies of the Indian Pines dataset for w

分类精度(%)	$w=3$	$w=5$	$w=7$	$w=9$	$w=11$	$w=13$	$w=15$
OA	97.00 \pm 0.26	97.62 \pm 0.36	97.72 \pm 0.35	97.88 \pm 0.26	97.52 \pm 0.29	97.85 \pm 0.28	97.71 \pm 0.43
AA	95.06 \pm 1.11	95.78 \pm 1.62	95.59 \pm 1.19	95.59 \pm 1.03	96.03 \pm 1.44	95.90 \pm 1.41	95.80 \pm 0.89
Kappa	96.58 \pm 0.30	97.29 \pm 0.41	97.40 \pm 0.39	97.58 \pm 0.29	97.17 \pm 0.30	97.55 \pm 0.37	97.39 \pm 0.49

通过交叉验证法分析了 k 值对高光谱图像分类精度的影响。选取 Indian Pines 数据集,设置 k 值从 1 到 40,试验结果如图 8 所示。随着 k 值增加,OA 总体呈下降趋势,这是因为“空-谱”联合后的同类样本间已经具有很高的相似性。如果 k 值选取过大,并且待分类样本又属于训练集中样本数较少的类,在选择较大 k 值时,就可能将不相关数据包含进来,导致分类精确度降低,同时增大计算成本。因此,在确保较高分类精确度同时, k 值选取尽可能小。当 $k=1$ 时,本算法在 Indian Pines 数据集上 OA 最高。后续实验对 Pavia University 数据集与 Salinas 数据集采用同样的方法进行参数选择与设置。

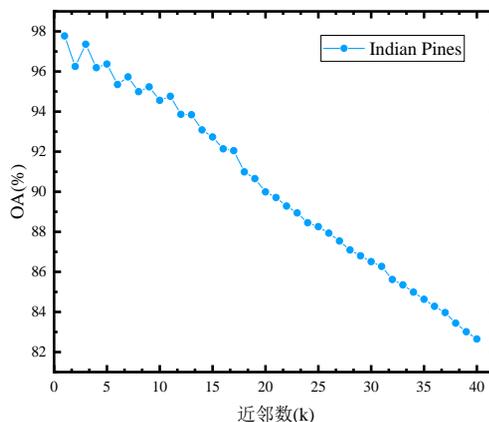


图8 Indian Pines 数据集近邻数 k 对分类精度的影响

Fig. 8 Influence on classification accuracies of the Indian Pines dataset for k

2.3 实验结果与分析

为了进一步验证本算法的有效性,采用 Pavia University、Indian Pines、Salinas 三种数据集进行验证,且采用了 KNN、基于径向基核函数的支持向量机(RBF-SVM)、核联合正交匹配追踪(Kernel Simultaneous Orthogonal Matching Pursuit, KSOMP)三种经典分类算法进行对比试验(表 6),每种分类方法所用的训练样本数完全相同。

Pavia University 数据集中,从每类地物随机选

取了10%样本作为训练集,其余样本用作测试集。训练样本总数4278个,包含10次实验结果的平均值和标准偏差,OA、AA、Kappa作为分类精度评价指标。四组对比试验中,KNN算法采用原始纯光谱特征进行分类, k 值设置为8,达到最优效果。SVM算法采用LibSVM^[27]工具箱实现分类,RBF-SVM^[28]为基于高斯核函数RBF的分类算法,内部主要函数均为C语言实现,惩罚因子 c 和参数 $gamma$ 由十次交叉验证法获取。KSOMP算法中的稀疏度 k_0 与径向基宽度 γ 分别设置为25和4。LBP-SSKNN方法的光谱特征采用降维后的前三个主成分,即 $p=3$, s 与 r 分别取8和1, w 设置为7像素 \times 7像素,最近邻分类器 k 值取1,此时分类效果最优,OA、AA、Kappa分别达到了99.15%、97.94%、98.87%。对于大多数地物,OA、AA、Kappa系数都有明显提高。同时,LBP-SSKNN与KSOMP都比传统仅使用纯光谱的SVM与KNN分类效果好,主要因为在利用光谱特征同时,也更好地利用了高光谱图像中蕴含的丰富空间信息,改善算法的鉴别能力,提升分类精度。在OA、

AA、Kappa方面,LBP-SSKNN比KSOMP分别高出5.04%、7.37%与6.79%,并且标准差更小,分类精度更稳定。

图9为Pavia University数据集在四种分类方法下的分类结果。可以看出,LBP-SSKNN得到的分类图总体错分点更少,大部分区域完整,更接近于原始地物分布图。另外,在最容易被错分的地物类型上表现更好,如Bare Soil、Gravel等,表明不仅对大部分的地物分类精度都有显著提升,而且对易错分地物类型也能保持良好的分类精度。Indian Pines数据集不同类别具有很高的相似性,且个别样本量较少,使得对该数据集的分类更加困难。对Indian Pines数据集同样采用从各类地物样本中随机选取10%作为训练样本,其余作为测试样本,训练样本总数为1027个,结果取10次实验的平均值。由表7和图10所示,根据实验结果,对于Indian Pines数据集,将 p 设置为4, w 为9像素 \times 9像素, r 与 s 分别设置为1和8,此时能得到最好的分类效果,其他算法均为最优时分类精度。

表6 Pavia University数据集四种分类方法下的精度(%)

Table 6 Classification accuracies using the four methods based on the Pavia University dataset

序号	地物种类	训练样本数	测试样本数	KNN	RBF-SVM	KSOMP	LBP-SSKNN
1	Asphalt	663	5 968	80.22 \pm 0.70	91.42 \pm 0.40	87.67 \pm 0.56	99.90 \pm 0.11
2	Meadows	1 865	16 784	99.18 \pm 0.09	96.54 \pm 0.51	99.99 \pm 0.00	99.97 \pm 0.03
3	Gravel	210	1 889	64.28 \pm 2.37	71.52 \pm 2.39	89.63 \pm 3.12	99.55 \pm 0.33
4	Trees	306	2 758	79.06 \pm 0.77	92.88 \pm 0.50	94.04 \pm 1.13	92.67 \pm 1.49
5	Painted metal sheets	135	1 211	99.21 \pm 0.27	99.52 \pm 0.26	100.00 \pm 0.00	99.33 \pm 0.33
6	Bare Soil	503	4 526	40.62 \pm 1.33	75.57 \pm 1.68	85.62 \pm 2.40	100.00 \pm 0.00
7	Bitumen	133	1 197	79.39 \pm 1.51	80.07 \pm 1.55	89.49 \pm 1.09	99.89 \pm 0.15
8	Self-Blocking Bricks	368	3 314	83.84 \pm 2.01	86.06 \pm 1.24	94.78 \pm 0.77	99.48 \pm 0.62
9	Shadows	95	852	92.96 \pm 0.86	97.52 \pm 1.31	73.95 \pm 0.44	90.63 \pm 1.29
OA				84.13 \pm 0.32	90.49 \pm 0.19	94.11 \pm 0.38	99.15 \pm 0.15
AA				79.86 \pm 0.40	87.90 \pm 0.45	90.57 \pm 0.41	97.94 \pm 0.27
Kappa				78.24 \pm 0.46	87.30 \pm 0.26	92.08 \pm 0.52	98.87 \pm 0.20

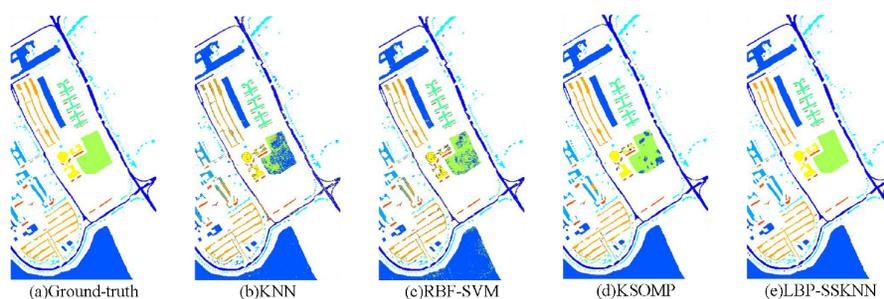


图9 Pavia University数据集四种分类方法划分结果

Fig. 9 Classification maps using the four methods for the Pavia University dataset

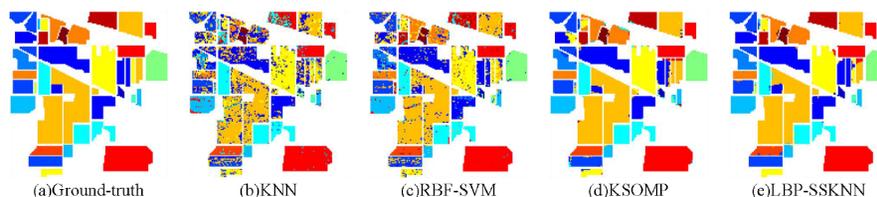


图 10 Indian Pines 数据集四种分类方法划分结果

Fig. 10 Classification maps using the four methods for the Indian Pines dataset

LBP-SSKNN 在 OA、AA、Kappa 上同其他分类算法的比较结果如表 7 所示。虽然 KNN 与 RBF-SVM 算法能够有效区分如 Wheat, Woods 等光谱差异大的地物,然而由于未考虑高光谱数据的空间信息,在 Alfalfa、Corn、Oats 等样本量少且易被错分的地物分类上,基于纯光谱特征的 KNN 与 RBF-SVM 分类效果并不理想。KNN 在 Alfalfa 地物类型的 OA 低于 25%,RBF-SVM 低于 20%。相比之下,KSOMP 对 Alfalfa 的 OA 达到了 86.77%,LBP-SSKNN 提升到了 93% 以上,且在其他样本量少、易错分的相似光谱地物分类中,如 Corn、Grass-pasture-mowed、Stone-Steel-Towers, OA 均达到了 92% 以上。在 Indian Pines 高光谱数据集上,KNN 与 RBF-KNN 的 OA 分别为 68.42% 和 81.26%,KSOMP 的 OA、AA、Kappa 分别达到了 93.94%、85.19%、93.08%,LBP-SSKNN

的 OA、AA 分别提升了 3.94%、10.41%,Kappa 系数提升了 4.5%,且 OA、Kappa 标准偏差分别为 0.26 与 0.29。总体来说,LBP-SSKNN 分类结果更稳定。

在 Salinas 数据集中,从每类地物中随机选取 2% 作为训练样本,其余作为测试样本,测试样本量为 1077 个,结果取 10 次实验结果平均值(表 8 和图 11)。将 Salinas 数据集前 2 个主成分作为光谱信息, w 用 17 像素 \times 17 像素窗口, r 与 s 分别设置为 1 和 8,此时达到最优分类效果,其他算法均为最优状态下的分类精度。可以看出,同一地物类别分布的集中,且每类地物样本量较为充足,因此该数据集相对易于区分不同地物。KNN 与 RBF-SVM 的 OA 分别为 87.02%、91.43%,但对于容易被错分的 Vineyard_untrained 分类效果不理想,噪声点较多,OA 仅为 53.64% 和 67.78%。相比之下,KSOMP 有效消除

表 7 Indian Pines 数据集四种方法下的分类精度(%)

Table 7 Classification accuracies using the four methods for the Indian Pines dataset

序号	地物种类	训练样本数	测试样本数	KNN	RBF-SVM	KSOMP	LBP-SSKNN
1	Alfalfa	5	41	24.88 \pm 10.22	17.32 \pm 14.69	86.77 \pm 7.11	93.41 \pm 4.32
2	Corn-notill	143	1285	59.64 \pm 4.23	75.61 \pm 2.16	88.44 \pm 1.72	97.95 \pm 0.41
3	Corn-mintill	83	747	45.02 \pm 2.68	69.97 \pm 2.56	90.80 \pm 3.08	96.28 \pm 1.51
4	Corn	24	213	24.88 \pm 8.06	46.95 \pm 6.17	91.74 \pm 4.19	95.45 \pm 2.17
5	Grass-pasture	48	435	80.74 \pm 1.91	90.69 \pm 2.29	93.70 \pm 0.10	97.52 \pm 0.77
6	Grass-trees	73	657	97.35 \pm 1.57	95.25 \pm 1.44	99.06 \pm 0.39	97.84 \pm 1.58
7	Grass-pasture-mowed	3	25	66.40 \pm 9.12	21.20 \pm 15.02	47.29 \pm 9.01	94.00 \pm 4.73
8	Hay-windrowed	48	430	97.28 \pm 2.15	99.14 \pm 0.57	99.95 \pm 0.09	99.60 \pm 0.56
9	Oats	2	18	16.60 \pm 5.00	6.11 \pm 0.38	0.00 \pm 0.00	78.33 \pm 1.78
10	Soybean-notill	97	975	68.61 \pm 3.06	69.51 \pm 3.05	89.65 \pm 3.16	96.88 \pm 1.48
11	Soybean-mintill	246	2209	73.13 \pm 1.94	86.52 \pm 1.03	96.54 \pm 0.78	98.90 \pm 0.49
12	Soybean-clean	59	534	27.08 \pm 4.68	71.89 \pm 4.92	93.61 \pm 2.39	96.72 \pm 1.51
13	Wheat	21	184	91.74 \pm 2.02	96.79 \pm 1.51	99.51 \pm 0.45	95.82 \pm 2.98
14	Woods	127	1138	93.50 \pm 0.72	96.78 \pm 1.36	99.32 \pm 0.57	99.09 \pm 0.37
15	Buildings-Grass-Trees-Drives	39	347	15.53 \pm 1.86	52.62 \pm 4.44	87.92 \pm 7.58	98.96 \pm 0.61
16	Stone-Steel-Towers	9	84	83.93 \pm 2.08	88.81 \pm 4.03	97.63 \pm 1.30	92.74 \pm 4.90
OA				68.42 \pm 0.60	81.27 \pm 0.56	93.94 \pm 0.46	97.88 \pm 0.26
AA				59.46 \pm 1.20	67.82 \pm 0.92	85.19 \pm 0.60	95.60 \pm 1.03
Kappa				63.71 \pm 0.71	78.48 \pm 0.66	93.08 \pm 0.53	97.58 \pm 0.29

了类间相似的误分类,OA 达到 84% 以上。LBP-SSKNN 对 Salinas 数据集大部分地物类型 OA 均高于 90%,且相比于 KSMOP,对 Vinyard_untrained 的识别,分类精度提高了 14.37%,OA 与 Kappa 系数分别提高了 2.06% 和 2.23%。在 Salinas 数据集上,LBP-SSKNN 分类精度标准偏差最小,说明不但能得到理想的分类精度,而且分类效果稳定。

此外,训练样本的数量也对高光谱图像分类精度有重要影响,而实际应用中训练样本数量往往有限。为了进一步验证本方法的有效性,对不同数据集随机抽取不同比例的训练样本进行测试,设置了

1%、2%、5%、10%、15%、20% 五种比例,图 12~14 分别显示了三种数据集在不同训练样本比例下的分类精度,每次实验取十次分类结果的平均值。总体来看,随着训练样本量的增加,不同算法的分类精度也不断提升,且逐渐趋于平稳。越多的训练样本量则可以提供更多具有代表性的先验信息,分类效果更好。在不同训练样本量下,联合光谱与空间信息的 KSOMP 和 LBP-SSKNN,大部分情况下的 OA、AA、Kappa 都高于仅使用光谱信息的 KNN 和 SVM。当训练样本比例为 1% 时,对于 Pavia University 数据集,KSOMP 与 LBP-SSKNN 都具有 85% 以上的 OA、

表 8 Salinas 数据集四种方法下的分类精度(%)

Table 8 Classification accuracies using the four methods for the Salinas dataset

序号	地物种类	训练样本数	测试样本数	KNN	RBF-SVM	KSOMP	LBP-SSKNN
1	Brocoli_green_weeds_1	40	1 969	97.17±0.48	98.60±1.15	99.91±0.12	99.81±0.32
2	Brocoli_green_weeds_2	69	3 357	98.22±0.33	99.01±0.38	99.97±0.06	99.96±0.05
3	Fallow	40	1 936	93.39±2.08	93.63±3.52	96.99±1.76	99.91±0.12
4	Fallow_rough_plow	28	1 366	98.87±0.11	98.59±0.83	99.45±0.34	93.11±2.17
5	Fallow_smooth	54	2 624	94.55±1.32	97.74±1.05	98.29±1.27	94.65±1.47
6	Stubble	79	3 880	99.52±0.09	99.44±0.23	100.00±0.00	97.32±0.95
7	Celery	72	3 507	99.21±0.08	99.35±0.25	99.05±0.42	98.56±0.75
8	Grapes_untrained	225	11 046	82.36±2.14	87.44±1.17	95.08±1.15	99.79±0.17
9	Soil_vinyard_develop	124	6 079	97.27±0.19	98.64±0.63	99.97±0.32	99.99±0.03
10	Corn_senesced_green_weeds	66	3 212	84.14±2.27	92.68±1.76	96.77±0.83	98.69±0.52
11	Lettuce_romaine_4wk	21	1 047	92.00±2.00	94.16±4.44	94.29±6.79	95.57±4.24
12	Lettuce_romaine_5wk	39	1 888	99.99±0.02	99.55±0.49	99.99±0.02	96.72±1.18
13	Lettuce_romaine_6wk	18	898	97.76±0.28	97.08±2.08	98.87±0.61	92.49±2.90
14	Lettuce_romaine_7wk	21	1 049	88.52±2.58	92.83±2.44	99.42±0.29	91.04±4.60
15	Vinyard_untrained	145	7 123	53.64±2.80	67.78±2.70	84.44±3.40	98.81±0.43
16	Vinyard_vertical_trellis	36	1 771	84.11±3.12	96.35±1.62	99.09±0.52	99.99±0.02
OA				87.02±0.26	91.43±0.39	96.23±0.40	98.46±0.15
AA				91.30±0.40	94.56±0.40	97.60±0.38	97.28±0.21
Kappa				85.51±0.29	90.44±0.44	95.81±0.45	98.29±0.17

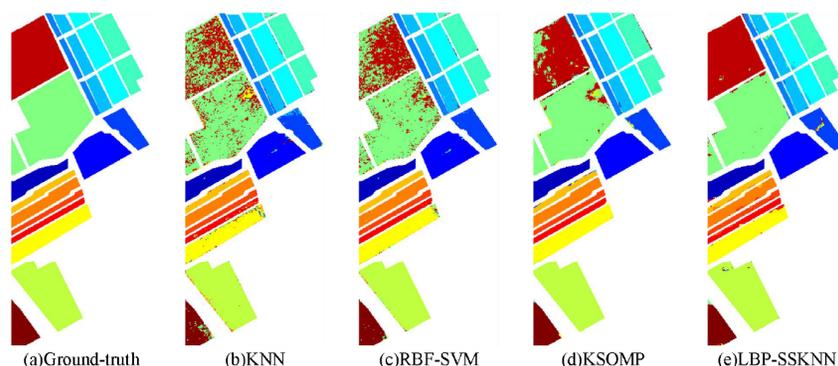


图 11 Salinas 数据集四种分类方法划分结果

Fig. 11 Classification maps using the four methods for the Salinas dataset

AA、Kappa系数,而在Indian Pines和Salinas数据集上,LBP-SSKNN有最高的OA和AA,并且在三种数据集上,Kappa系数均在80%以上。文献[16]所提出的一种新颖的SVMMRF-NE算法在Indian Pines数据集上的OA、Kappa系数分别为92.05%、90.93%,所用样本总量为695个,本算法取5%的训练样本量为512个,此时OA和Kappa分别达到95.24%与94.58%。文献[18]提出的WSSD-KNN

算法在Indian Pines数据集上的OA、AA、Kappa分别达到91.72、92.49%、90.55%,本算法则分别提升了6.16%、3.11%、7.03%。当训练样本为6%时,文献[18]对Pavia University数据集OA为96.56%,本算法在5%训练样本下的OA为98.11%。结果表明,相比于其他分类算法,当训练样本量较小时,LBP-SSKNN具有更高的精度,且当训练样本量足够大时,则具有最高的分类精度。

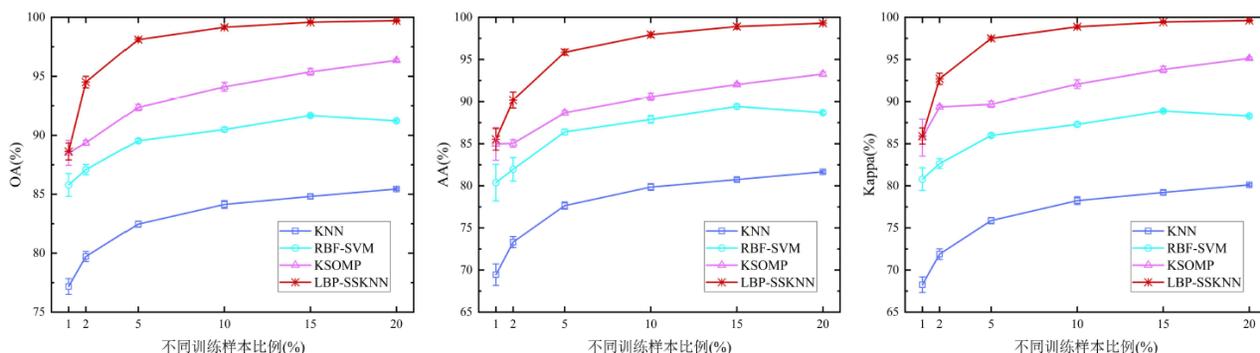


图12 Pavia University数据集在不同训练样本比例的分类精度

Fig. 12 Classification accuracies under different training samples for the Pavia University dataset

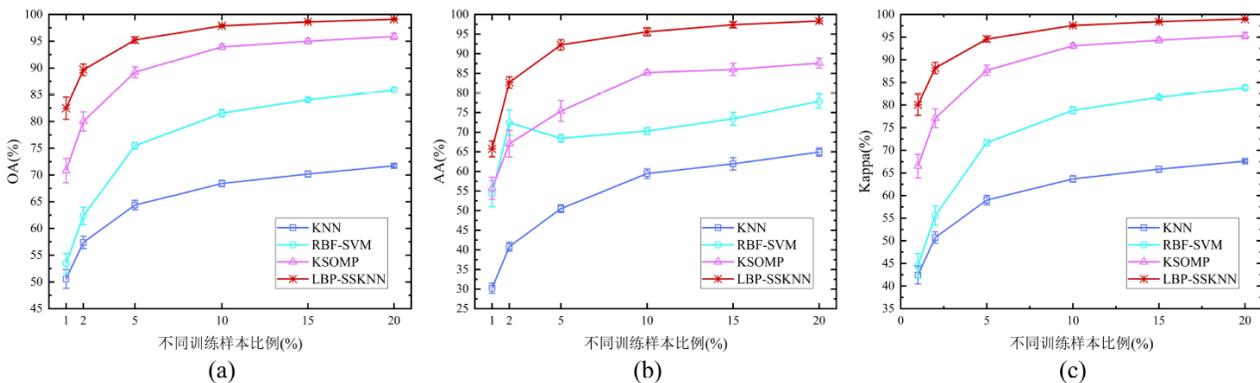


图13 Indian Pines数据集在不同训练样本比例的分类精度

Fig. 13 Classification accuracies under different training samples for the Indian Pines dataset

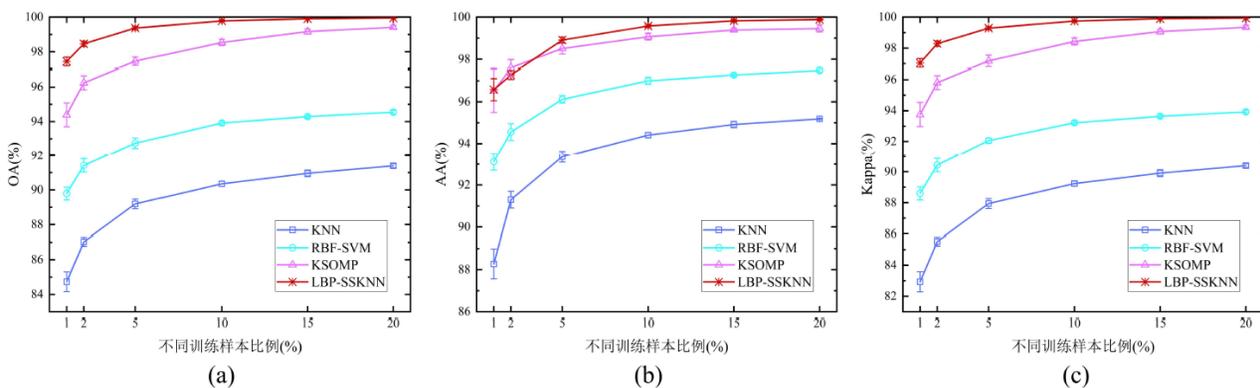


图14 Salinas数据集在不同训练样本比例的分类精度

Fig. 14 Classification accuracies under different training samples for the Salinas dataset

3 结论

针对高光谱图像分类中训练样本有限、高相关性、冗余量大等特点,提出了一种联合LBP和KNN的高光谱图像分类方法LBP-SSKNN。通过LBP提取降维后的高光谱图像纹理信息,以向量堆叠的方式联合光谱与空间特征,再将联合后的空-谱信息输入KNN分类器分类,并在三种公开高光谱数据集上验证了分类性能。实验结果表明,LBP-SSKNN通过联合光谱与空间特征,增强了同类像元的相似性,同时增大了不同像元的相异性,与KNN、SVM、KSOMP分类结果和部分文献算法相比,总体分类精度(OA)、平均分类精度(AA)、Kappa系数均有不同程度的提高。此外,当训练样本数量较小时,也能保持较高的分类精度和稳定性。本算法采用的KNN分类器,具有可调参数少、低复杂度等优点,更加适用于实际应用场景。后续的研究方向可考虑如何在保持低计算复杂度同时,除了融合空间与光谱特征外,可联合颜色、形状、形态等新的互补特征,进一步提高高光谱图像分类精度和效率,以应对更加复杂的应用场景。

References

- [1] Chen C, Wei L, Hong J S, et al. Spectral-spatial classification of hyperspectral image based on Kernel Extreme Learning Machine[J]. *Remote Sensing*, 2014, **6**(6):5795-5814.
- [2] Yu C Y, Lee L, Bai X, et al. Band-specified virtual dimensionality for band selection: An orthogonal subspace projection approach [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, **56**(5):2822-2832.
- [3] Wang Q, Zhang F H, Li X L. Optimal clustering framework for hyperspectral band selection [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2018, **56**(10):5910-5922.
- [4] Abdi H, Williams L J. Principal component analysis [J]. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2010, **2**(4):433-459.
- [5] Luo F L, Du B, Zhang L P, et al. Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image [J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 2019, **49**(7):2406-2419.
- [6] Zhang S C, Li X L, Zong M, et al. Efficient kNN classification with different numbers of nearest neighbors [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2018, **29**(5):1774-1785.
- [7] Tan Kun, Du Pei-Jun. Hyperspectral remote sensing image classification based on support vector machine [J]. *J. Infrared Millim. Waves.* (谭琨, 杜培军. 基于支持向量机的高光谱遥感图像分类. *红外与毫米波学报*), 2008, **27**(2):123-128.
- [8] Liu Y, Lu S, Lu X T, et al. Classification of urban hyperspectral remote sensing imagery based on optimized spectral angle mapping [J]. *Journal of The Indian Society of Remote Sensing*, 2019, **47**(2):289-294.
- [9] Haut J M, Paoletti M E, Plaza J, et al. Fast dimensionality reduction and classification of hyperspectral images with extreme learning machines [J]. *Journal of Real Time Image Processing*, 2018, **15**(3):439-462.
- [10] Rajan S, Ghosh J, Crawford M M. An active learning approach to hyperspectral data classification [J]. *IEEE Transactions on Geoscience & Remote Sensing*, 2008, **46**(4):1231-1242.
- [11] Li J, Khodadadzadeh M, Plaza A, et al. A discontinuity preserving relaxation scheme for spectral - spatial hyperspectral image classification [J]. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 2016, **9**(2):625-639.
- [12] Cao X, Zhou F, Xu L, et al. Hyperspectral image classification with Markov random fields and a convolutional neural network [J]. *IEEE Transactions on Image Processing*, 2018, **27**(5):2354-2367.
- [13] Jin Rui, Li Xiao-Xian, Yan Yi-Yun, et al. Detection method of multi-target recognition of potato based on fusion of hyperspectral imaging and spectral information [J]. *Transactions of the Chinese Society of Agricultural Engineering.* (金瑞, 李小显, 颜伊芸, 等. 基于高光谱图像和光谱信息融合的马铃薯多指标检测方法, *农业工程学报*), 2015, **31**(16):258 - 263.
- [14] Zhu Z X, Jia S, He S, et al. Three-dimensional Gabor feature extraction for hyperspectral imagery classification using a memetic framework [J]. *Information Sciences*, 2015, **298**(298):274-287.
- [15] Li W, Du Q. Gabor-filtering-based nearest regularized subspace for hyperspectral image classification [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2014, **7**(4):1012-1022.
- [16] Tarabalka Y, Fauvel M, Chanussot J, et al. SVM- and MRF-based method for accurate classification of hyperspectral images [J]. *IEEE Geoscience and Remote Sensing Letters*, 2010, **7**(4):736-740.
- [17] Zhang B, Li S S, Jia X P, et al. Adaptive Markov random field approach for classification of hyperspectral imagery [J]. *IEEE Geoscience & Remote Sensing Letters*, 2011, **8**(5):973-977.
- [18] Huang Hong, Zheng Xin-lei. Hyperspectral image classification with combination of weighted spatial-spectral and KNN [J]. *Optics and Precision Engineering.* (黄鸿, 郑新磊. 加权空-谱与最近邻分类器相结合的高光谱图像分类. *光学精密工程*), 2016, **24**(4):873-881.
- [19] Chen Y, Nasrabadi N M, Tran T D. Hyperspectral image classification via kernel sparse representation [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2013, **51**(1):217-231.
- [20] Tropp J A, Gilbert A C, Strauss M, et al. Algorithms for simultaneous sparse approximation: part I: Greedy pursuit [J]. *Signal Processing*, 2006, **86**(3):572-588.
- [21] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2002, **24**(7):971-987.
- [22] Pearson K. LIII. On lines and planes of closest fit to sys-

- tems of points in space[J]. *Philosophical Magazine, Series 6*, 1901, **2**(11):559-572.
- [23] Cover T M, Hart, P E. Nearest neighbor pattern classification[J]. *Information Theory*, 1967, **13**, 21-27.
- [24] Townsend J T. Theoretical analysis of an alphabetic confusion matrix [J]. *Attention Perception & Psychophysics*, 1971, **9**(1):40-50.
- [25] Foody G M. Status of land cover classification accuracy assessment [J]. *Remote Sensing of Environment*, 2002, **80** (1):185-201.
- [26] Wang Xiao-Fei, Zhang Jun-Ping, Zhang Ye. Unmixing algorithm of hyperspectral image [J]. *J. Infrared Millim. Waves*. (王晓飞, 张钧萍, 张晔. 高光谱图像混合像元分解算法. *红外与毫米波学报*), 2010, **29**(3):210-215, 229.
- [27] Ruske S, Topping D, Foot V E, *et al.* Evaluation of machine learning algorithms for classification of primary biological aerosol using a new UV-LIF spectrometer [J]. *Atmospheric Measurement Techniques*, 2016, **10** (2) : 695-708.
- [28] Man C D, Nguyen T T, Bui H Q, *et al.* Improvement of land-cover classification over frequently cloud-covered areas using Landsat 8 time-series composites and an ensemble of supervised classifiers [J]. *International Journal of Remote Sensing*, 2018, **39**(4):1243-1255.