

联合多尺度多特征的高分遥感图像场景分类

黄 鸿,徐科杰,石光耀

(重庆大学光电技术及系统教育部重点实验室,重庆 400044)

摘 要: 高分辨率遥感图像地物信息丰富,但场景构成复杂,目前基于手工设计的特征提取方法不能满足复杂场景分类的需求,而非监督特征学习方法尽管能够挖掘局部图像块的本征结构,但单一尺度及尺度的特征难以有效表达实际应用中复杂遥感场景特性,导致分类性能受限. 针对此问题,本文提出了一种基于多尺度多特征的遥感场景分类方法. 该算法首先设计了一种改进的谱聚类非监督特征(iUFL-SC)以有效表征图像块的本征结构,然后通过密集采样提取每幅遥感场景的 iUFL-SC、LBP、SIFT 等三种多尺度局部图像块特征,并通过视觉词袋模型(BoVW)获得场景的中层特征表达,以实现更为准确详实的特征描述,最后基于直方图交叉核的支持向量机(HIKSVM)进行分类. 在 UC Merced 数据集以及 WHU-RS19 数据集上的实验结果表明本文方法可对遥感场景进行鉴别特征提取,有效提高分类性能.

关键词: 遥感; 高分辨率影像; 场景分类; 非监督特征; 特征融合; 视觉词袋模型

中图分类号: TP751.1 **文献标识码:** A **文章编号:** 0372-2112 (2020)09-1824-10

电子学报 URL: <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2020.09.021

Scene Classification of High-Resolution Remote Sensing Image by Multi-scale and Multi-feature Fusion

HUANG Hong, XU Ke-jie, SHI Guang-yao

(Key Laboratory of Optoelectronic Technique System of the Ministry of Education, Chongqing University, Chongqing 400044, China)

Abstract: High resolution image possesses abundant information of ground objects. The hand-crafted features cannot meet the demand of complex scene classification due to complex scene distribution, while the unsupervised feature learning method can exploit the intrinsic structure of image patches to obtain effective discriminating features. However, single feature with a scale is difficult to represent the characteristics of complex scenes in practical applications, which restricts classification performance. To solve this problem, this paper proposed a new method based on multi-scale and multi-feature fusion (MMF) for remote sensing scene classification. At first, an improved unsupervised feature learning via spectral clustering (iUFL-SC) is designed to effectively reveal the intrinsic structure of image patches, and then the iUFL-SC, LBP, and SIFT features of image patches are extracted by dense sampling in each image. After that, the middle-level features of each scene are obtained through bag of visual words (BoVW) model for effective feature description. Finally, the fused features are classified by histogram intersection kernel SVM. Experimental results on two public data sets indicate that MMF can extract discriminant features of remote sensing image and subsequently improve the classification performance.

Key words: remote sensing; high resolution images; scene classification; unsupervised features; feature fusion; bag of visual words

1 引言

随着中国高分辨率对地观测系统重大专项的启动实施,遥感图像获取实现了跨越式发展,已经形成高空间分辨率、高时间分辨率和高光谱分辨率的对地观测系统. 目前,通过高分系列卫星采集的亚米级高分辨率遥感图像已经为土地资源规划利用、公共安全监测、防灾减灾、环

境保护等重要领域提供信息服务和数据支撑. 相比于中、低分辨率遥感图像,高分辨率遥感图像包含丰富而详尽的地物信息,可反映出准确的地物类型及其空间分布,但其地物空间分布复杂,同物异谱及异物同谱现象突出,给遥感图像场景分类带来了极大挑战^[1].

视觉词袋(Bag of Visual Words, BoVW)模型是一种经典的中层特征编码方法,因其简单高效的特性被广泛

收稿日期:2019-03-26;修回日期:2020-04-19;责任编辑:梅志强

基金项目:重庆市基础与前沿研究计划(No. cstc2018jcyjAX0093);重庆市留学人员回国创业创新支持计划(No. cx2019144);重庆市研究生科研创新项目(No. CYB19039);重庆市教委科学技术研究计划(No. KJZD-K201902501)

用于场景的中层表示^[2]. 在 BoVW 模型中,首先利用 K 均值聚类算法构造基于图像局部特征的视觉词汇字典;然后通过矢量化(Vector Quantization, VQ)将所有局部特征映射到相应的视觉词汇,并统计各个视觉词汇出现的频率,由此形成的直方图即为图像的特征表达^[3-6]. 在 BoVW 模型的框架下,尺度不变特征变换(Scale Invariant Feature Transform, SIFT)^[7]、颜色直方图和局部二值模式特征(Local Binary Pattern, LBP)^[8]等基于手工设计(hand-crafted)的低层次特征常常被用作局部特征,在一定程度上提升了场景分类效果. 由于深度学习方法具有强大的特征学习能力,近年来被广泛应用于特征提取及分类任务中^[9-12]. Penatti 等^[13]首次使用由自然图像预训练的深层神经网络来提取全局特征,取得了不错的效果. 然而,现有深度学习模型的参数大都是在自然图像大数据集训练得到,导致直接应用于遥感场景分类,其特征鉴别性能受限,且对该模型进行训练计算复杂度非常高. 与此同时,非监督特征学习(Unsupervised Feature Learning, UFL)方法可挖掘图像中丰富的边缘、形状等信息,进而自适应学习数据本身的鉴别特性,有效解决了遥感图像场景描述问题^[14,15]. 在一些图像分类应用中,UFL 特征学习方法甚至可以替代手工标注方法. 但是,目前大都采用单一特征、单一尺度的学习方法,导致其难以全面描述场景信息,且单一特征对于不同场景的适应性也受限.

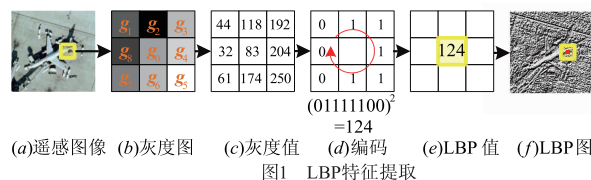
针对上述问题,本文在传统 BoVW 编码模型的基础上,提出了一种多尺度多特征融合(Multi-scale and Multi-feature Fusion, MMF)的高分辨率遥感图像场景分类算法. 该算法首先通过密集采样提取每幅场景的三种多尺度局部特征,包括改进的谱聚类非监督特征(improved Unsupervised Feature Learning via Spectral Clustering, iUFL-SC)、LBP、SIFT 等特征. 然后通过 BoVW 模型获得场景的中层特征表达,最后采用基于直方图交叉核的支持向量机(Histogram Intersection Kernel Support Vector Machine, HIKSVM)进行分类. 在 UC Merced 和 WHU-RS19 数据集上的实验结果表明,本文提出的算法相比其他相关算法具有更好的场景分类性能.

2 相关方法

2.1 LBP 特征

LBP 特征是一种用于表征图像局部纹理结构的描述子,运算速度快、描述能力强,且具有旋转不变性和灰度不变性等特点. 在 3×3 的窗口内,LBP 算子以窗口中心像素的灰度为阈值,通过比较其周围 8 个像素的灰度进行编码. 若邻域点像素灰度值 g_i 大于中心像素灰度值 g_0 ,则该像素点的位置被编码为 1,否则为 0. 以左上角为起点,将所有编码值按顺时针连接可构成一个八

位的二进制数. 该二进制数换算的十进制数即为中心位置新的像素值. 具体如图 1 所示.

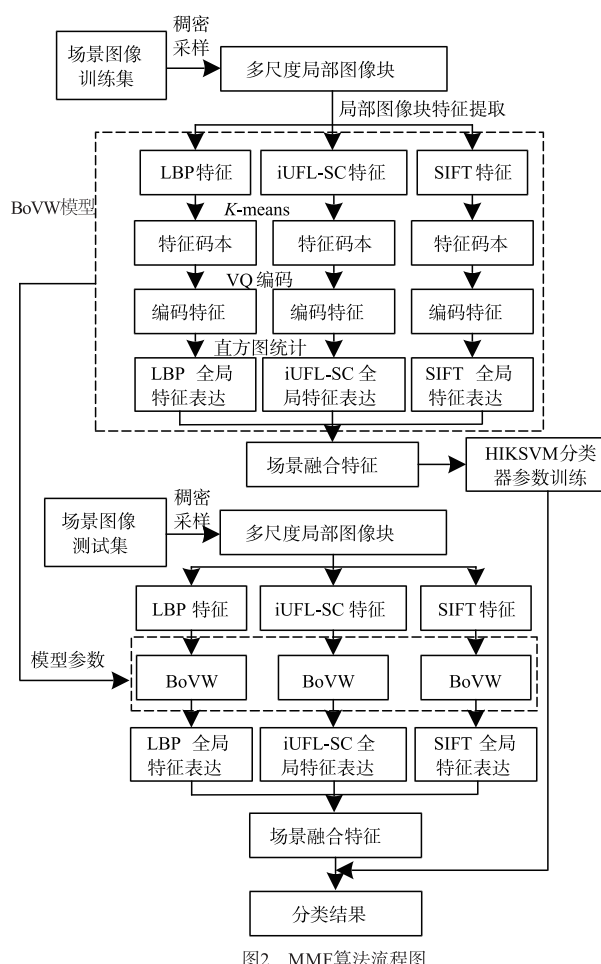


2.2 SIFT 特征

SIFT 特征是一种常见的局部特征描述符,具有旋转、缩放、平移不变性,在改变图像的亮度或部分遮挡,依然具有较好的特征描述效果. 经典的 SIFT 描述子首先找到特征点位置,通过在 4×4 空间网格上计算的梯度直方图来描述图像块,并将梯度量化到 8 个方向区间,最终得到 128 维的特征向量.

3 本文算法

本文首先设计了一种 iUFL-SC 特征以有效表征图像块的本征结构,自适应学习数据本身的鉴别特性. 然后,通过密集采样的方式采集局部图像块以提取每幅遥感场景的 iUFL-SC、LBP、SIFT 等三种多尺度局部特



征,并通过视觉词袋模型(BoVW)获得场景的中层特征表达,以实现更为准确详实的特征描述.最后,基于直方图交叉核的支持向量机(HKSVM)被用于分类.本文方法的总流程如图2所示.

3.1 iUFL-SC 特征提取

谱聚类非监督特征(Unsupervised Feature Learning via Spectral Clustering, UFL-SC)是一种基于非监督特征学习的方法,通过谱聚类算法有效挖掘局部图像块的本征结构,从而自适应的“学习”鉴别特征^[16].但遥感场景中

地物目标具有几何形状、尺度等差异性大的特点,若仅采用单一尺度的图像块,“学习”能力存在一定限制,且UFL-SC在非线性编码时仅以欧式距离作为数据划分的度量方式,易受空间分布影响,难以反映出数据划分的真实情况.相比于UFL-SC方法,本文提出的iUFL-SC采用多尺度图像块进行离线“学习”,可以得到丰富详实的字典及更为有效的投影矩阵.而且,在非线性编码过程中,融合欧式距离与角度两种度量方式,能有效的实现同类数据的真实划分.其具体步骤如图3所示.

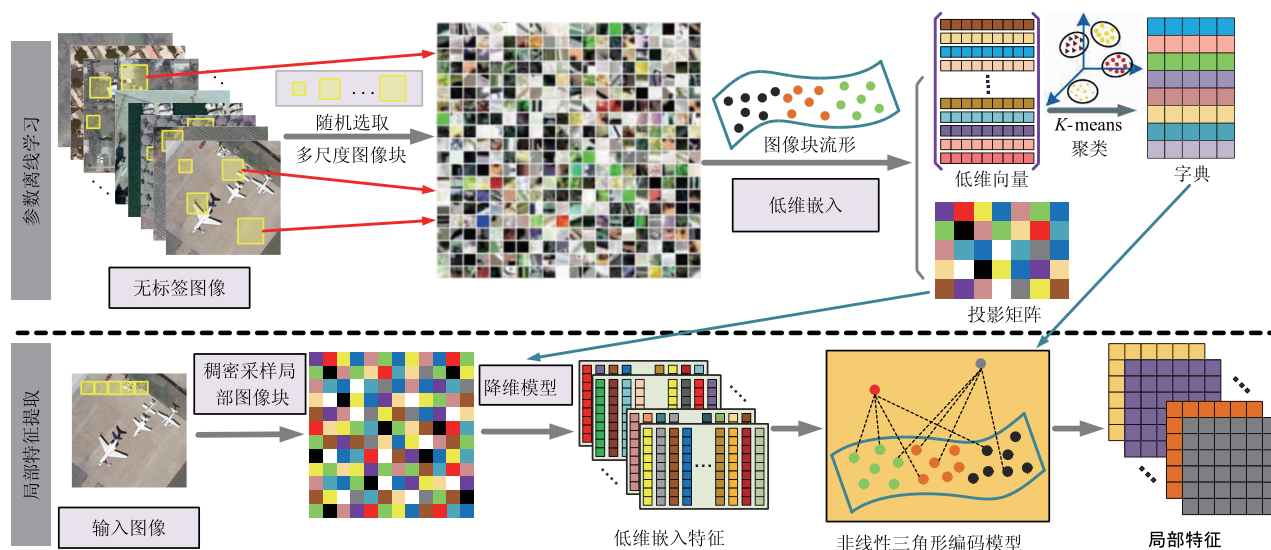


图3 iUFL-SC局部特征提取

该算法首先离线“学习”参数,训练得到投影矩阵 M 和字典 D_{ic} ,具体步骤如下.

步骤1 对于高分辨率遥感图像场景数据集 S ,在每一幅场景中随机采样一定数量的不同尺寸的局部图像块,并调整至相同大小.通过这种方式,我们提取到的图像块包含更多的有用信息与细节.

步骤2 将调整尺寸后的所有图像块转化为高维向量,得到局部图像块向量表示 X .例如, $n_1 \times n_1$ 大小的RGB局部图像块,可以转化成 $3 \times n_1 \times n_1$ 维的向量.

步骤3 对局部图像块向量 X 进行归一化及ZCA白化等预处理,降低数据的冗余性.

步骤4 采用局部保持投影算法(Locality Preserving Projections, LPP)^[17]可以得到投影矩阵 M 及图像块向量的低维表示 Y ,其主要步骤包括:(a)构造邻接图 G ;(b)定义权重 W ,如式(1)所示.

$$w_{ij} = \begin{cases} e^{-\frac{\|x_i - x_j\|^2}{t}}, & x_i \text{ 和 } x_j \text{ 为近邻} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

式中, t 为热核参数, w_{ij} 为 x_i 和 x_j 的权值.(c)求解式(2)中目标函数.

$$\min M^T X L X^T M, \quad \text{s. t. } M^T X D X^T M = I \quad (2)$$

式中, $D = \text{diag}([d_{11}, d_{22}, \dots, d_{NN}])$, $d_{ii} = \sum_{j=1}^N w_{ij}$, L 为拉普拉斯矩阵 $L = D - W$.

步骤5 采用K-means算法在图像块低维表示 Y 上训练含有 K 个聚类中心的字典 D_{ic} .

步骤6 得到字典 D_{ic} ,即可对数据集的每一幅场景提取局部特征,具体步骤如下:

(1)以 p 为采样间隔,稠密采样固定数量的局部图像块;

(2)对稠密采样得到的局部图像块进行步骤2和步骤3中的预处理,获取场景的局部图像块向量 X ;

(3)采用式(2)得到的投影矩阵 M ,通过式(3)计算可得低维向量 Y ;

$$Y = M^T X \quad (3)$$

(4)通过融合角度与欧式距离的度量方式实现非线性三角形编码,最终获取图像的局部特征,具体过程如下:

$$\phi_i = \max\left(0, \frac{1}{K} \sum_{k=1}^K \text{dist}_k - \text{dist}_i\right) \quad (4)$$

其中 dist 是一种融合度量方式,表示输入向量 y 与字典 D_{ic} 中每个聚类中心 d_j 的融合距离,计算方式如式(5):

$$\text{dist}_i = \frac{\alpha(D_{y_d} - \min(D_{y_{\cdot}}))}{\max(D_{y_{\cdot}}) - \min(D_{y_{\cdot}})} + \frac{(1 - \alpha)(\theta_{y_d} - \min(\theta_{y_{\cdot}}))}{\max(\theta_{y_{\cdot}}) - \min(\theta_{y_{\cdot}})} \quad (5)$$

式中, D_{y_d} 表示数据 y_i 与 d_j 的欧式距离; θ_{y_d} 表示数据 y_i 与 d_j 的余弦距离; α 为融合系数, 其值为 $0 \sim 1$; $D_{y_{\cdot}}$ 表示 y_i 与全部聚类中心的欧式距离; $\theta_{y_{\cdot}}$ 表示 y_i 与全部聚类中心的余弦距离; $\max(\cdot)$ 和 $\min(\cdot)$ 分别表示求最大和最小值。

3.2 BoVW 模型框架下的 MMF 分类方法

在高分遥感场景中, 地物类型种类多样、空间分布复杂, 所包含的地物目标在尺度、方向等属性上不尽相同, 因此语义类别相同的场景往往视觉差异较大。针对上述情况, 若仅编码单一尺寸的局部图像块, 则难以完全描述场景。因此, 本文将稠密采样 3 种尺度的图像块, 提取特征以形成多尺度联合局部特征, 具体如图 4 所示。其计算步骤如下:

(1) 以 p_1, p_2, p_3 为间隔提取不同尺度的局部图像块 ($n_1 \times n_1, n_2 \times n_2, n_3 \times n_3$, 其中 $n_1 < n_2 < n_3$), 且均调整到相同大小 $n_1 \times n_1$, 然后提取场景稠密局部特征 F_1, F_2, F_3 , 形成多尺度联合局部特征 $F = [F_1, F_2, F_3]$ 。

(2) 对于训练样本中的联合局部特征, 利用 K -means 计算得到关于场景类别 B 的特征码本 $C^{(B)}$; 对每个场景类别 B 均重复上述聚类操作, 并构造联合码本

$$C^J = [C^{(1)}, C^{(2)}, \dots, C^{(B)}].$$

(3) 对于每一幅场景, 先利用 VQ 编码将每幅场景对应的所有局部特征分别映射到联合码本中相似度最高的视觉词汇从而获取一组编码特征, 具体如式 (6) 所示。

$$\arg \min_H \|f_i - C^J h_i\|_2^2 \quad (6)$$

$$\text{s. t. } \|h_i\|_{l_0} = 1, \quad \|h_i\|_{l_1} = 1, \quad h_i \geq 0, \quad \forall i$$

式中, $H = [h_1, h_2, \dots, h_N]$ 表示局部特征 f_i 在字典 C^J 上投影而得到的 1 组稀疏系数。

(4) 统计局部特征映射到每个视觉词汇的频次, 最终得到的统计直方图即为全局特征, 其中视觉词汇的数量即为全局特征的维数。

(5) 对于场景图像, 重复 (1) 到 (4) 的过程, 可以分别获取 iUFL-SC 全局特征表达 H_{iUFL} 、LBP 全局特征表达 H_{LBP} 以及 SIFT 全局特征表达 H_{SIFT} , 并采用式 (7) 的行堆叠方式形成融合特征:

$$H_{\text{Fu}} = [H_{\text{iUFL}}, H_{\text{LBP}}, H_{\text{SIFT}}] \quad (7)$$

(6) 获得图像的融合特征后, 即可采用分类器进行场景分类。由于直方图交叉核是一种训练速度较快, 适合直方图特征相似性判别的 Mercer 核函数, 因此本文采用 HKSVM 分类器^[18] 进行分类, 直方图交叉核定义如下:

$$K_{\text{HK}}(H_m, H_n) = \sum_{k=1} \min(H_m(k), H_n(k)) \quad (8)$$

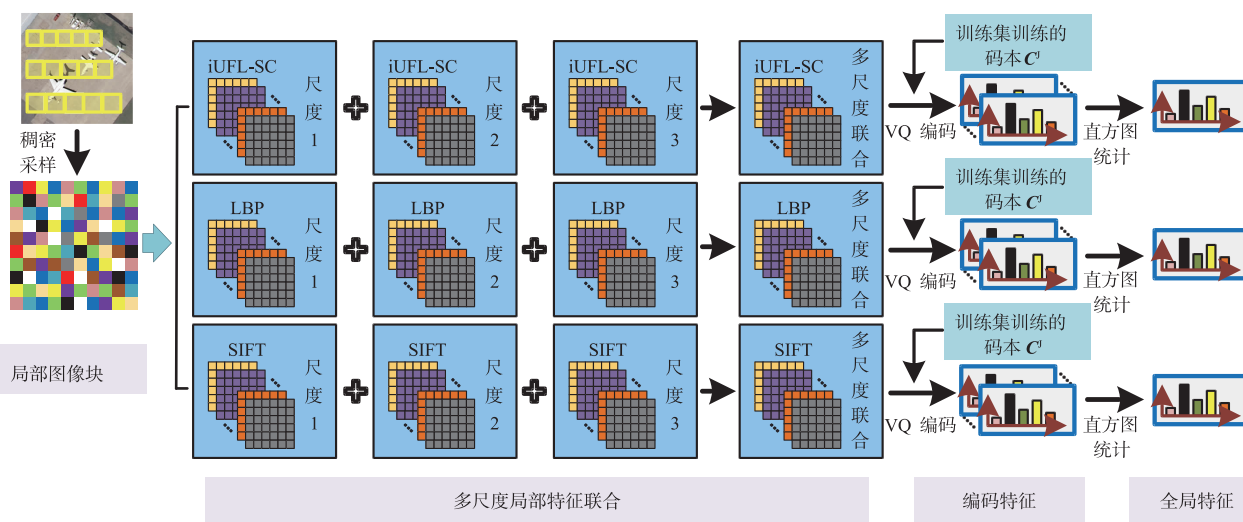


图4 BoVW模型框架下的多尺度局部特征编码

4 实验结果及分析

4.1 实验数据集

UC Merced 数据集^[2]: 该数据集的原始图像采集于美国地质调查局提供的国家地图。该数据集共包含 2100 张图片, 每幅场景大小为 256×256 , 分辨率约为 0.3m, 分为 21 类场景, 如农田、飞机、棒球场、建筑物等。UCM 数据集含有不同土地利用类型, 包括高密度住

宅区, 中密度住宅区和稀疏住宅区等高度相似的场景类别, 因此被广泛用于场景分类与检索相关的实验中。

图 5(a) 为该数据集部分场景图像。

WHU-RS19 数据集: 该数据集是武汉大学采集于 Google 卫星影像, 每幅场景大小为 600×600 , 分辨率约为 0.5m, 每类场景包含 50 张图片, 共 19 类场景, 主要包括机场、海滩、桥梁、商业区等, 具体如图 5(b) 所示。



图5 两个公开遥感图像数据集的示例图像

4.2 实验设置

在实验中,采用总体分类精度(Overall Accuracy, OA)作为场景分类性能评估指标.在每次实验中,按一定比例将数据集随机分成训练样本集和测试样本集,其中在UCM数据集中训练样本数与测试样本数之比为8:2,在WHU-RS19数据集中该比例为6:4.在相同实验条件下均重复进行40次实验,以降低因样本选取不同而造成的实验误差影响.

4.3 参数实验

本文算法包括融合系数 α 、视觉词汇个数(Number of Codebooks, N_{oc})等参数,为了评估它们对于分类性能

的影响,本文通过实验对其分析,并选取最佳参数.

4.3.1 融合系数 α

为研究融合系数 α 对iUFL-SC特征分类精度的影响,首先选用UCM数据集进行实验.实验中,从数据集的每幅场景中随机采样 10×10 、 30×30 、 50×50 的局部图像块各100个用于训练投影矩阵 M 和字典 D_{ic} .其中,采用LPP算法计算投影矩阵 M 时,将维数约简至150,并且令字典 D_{ic} 包含150个聚类中心.对于WHU数据集,为平衡计算时间与分类效果,每幅场景提取 20×20 、 25×25 、 30×30 三种大小的局部图像块各100个用于训练投影矩阵 M 和字典 D_{ic} .采用

LPP 算法将维数约简到 300 维,训练得到的字典包含 110 个聚类中心.此外,视觉词汇个数 N_{oc} 暂时分别设置为 50 和 100.

由图 6(a)可知,欧式距离与余弦距离间融合系数 α 的设定会影响场景的分类结果.在 UCM 数据集中,两者以 $\alpha=0.9$ 融合后取得了最佳效果;在 WHU 数据集中,两者以 0.1 作为系数融合后取得了最佳效果.这说明了在本文算法中,通过融合数据间余弦和欧氏距离进行非线性编码,在一定程度上减少了数据点受空间分布的影响,映射到同类聚类中心的概率更大,能更有效的提取图像的鉴别局部特征,最终提高分类精度.但在两个数据集中, α 的最优参数相差较明显.通过观察非线性三角形编码后的特征分布,不难发现 UCM 数据集中,欧式距离度量取得了更平稳的数据分布,而在 WHU 数据集中余弦距离度量取得更好的结果,具体见表 1.由于词袋模型是一种基于统计学原理的框架,将分布更合理的数据输入 BoVW 模型,可以获得相对较好的结果.由于两个数据集在场景类别、分辨率、尺寸等方面存在较大差异,其中 UCM 数据集中欧式距离度量占主导地位(α 值大),WHU 数据集中余弦距离度量占主导地位(α 值小).

表 1 两数据集中不同度量方式得到距离矩阵的标准差

数据集	标准差	
	欧式距离	余弦距离
UCM	0.1463	0.1818
WHU	0.2622	0.1711

4.3.2 视觉词汇个数 N_{oc}

为评估每类场景对应特征码本的视觉词汇个数 N_{oc} 对 MMF 算法分类效果的影响,在两个数据集分别对比了 1,2,5, ..., 300 共 17 种取值情况,从而获取最佳参数.在提取多尺度局部图像块的过程中由于数据集的差异性,因此选用不同的尺度.对于 UCM 数据集,场景大小为 256×256 且地物目标尺度变化较大,故局部图像块大小为 10×10 、 30×30 、 50×50 ,对应的采样间隔 p 选取 5、10、15.对于 WHU 数据集,场景大小为 600×600 ,选用过小的采样窗口将会产生大量图像块,导致计算复杂度的急剧上升.此外 WHU 数据集中,地物尺度变化相对较平缓,故局部图像块大小为 20×20 、 25×25 、 30×30 ,采样间隔 p 选取 5、12、15.

由图 6(b)可知,随着视觉词汇个数 N_{oc} 的增加,MMF 算法的分类精度在两个数据集中均呈现上升—稳定—下降的趋势.在 UCM 数据集实验中,参数 N_{oc} 取值范围为 1~5 时分类精度快速增长;而后 N_{oc} 值在 30~80 范围时,OA 保持平稳且精度较高;随着 N_{oc} 的继续增加,分类精度开始逐渐下降.在 WHU 数据集实验中, N_{oc} 取值为 1~

5 时分类精度增长较快,在为 40~120 时分类精度较高且逐渐平稳,而后开始略有下滑.由此可见,每类视觉词汇个数过少,对场景的描述能力有所欠缺,最终会造成分类效果不佳的现象;随着 N_{oc} 的增加,对场景的表示能力不断增强,分类精度也随之逐渐提高,但同时也增加了计算耗时;然而,当视觉词汇个数过多,得到的全局特征向量的维数过高,则会造成过拟合的现象.在 UCM 数据集中, N_{oc} 为 40 时,OA 达到了峰值 92.64%;在 WHU 数据集中, N_{oc} 为 100 时,OA 取得了 91.70% 的最好结果.综合考虑到算法分类精度以及运行效率,后文两个数据集实验中的取值分别为 40 和 100.两个数据集中 N_{oc} 取值差异较大,这是由于在场景类别数量类似(21 类和 19 类)的情况下,WHU 中图像(600×600)远大于 UCM 的图像(256×256),因此包含的信息更为丰富,需要更多数量的视觉词汇来提高场景的表示能力.

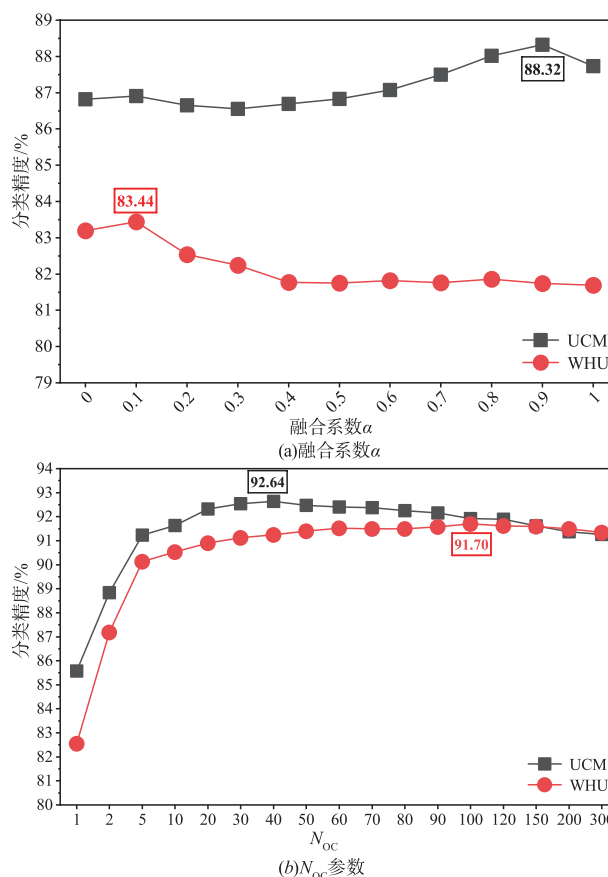


图 6 两个数据集中融合系数 α 和 N_{oc} 参数对场景分类性能的影响

4.4 UCM 数据集实验结果与分析

为了验证多尺度以及多特征对分类精度的影响,本文在 UCM 数据集上比较了 UFL-SC、iUFL-SC、多尺度 iUFL-SC、多特征融合 (Multi-feature Fusion, MF) 和本文提出的 MMF 等五种不同方法在不同场景类别下的分类性能,其实验结果如图 7 所示.

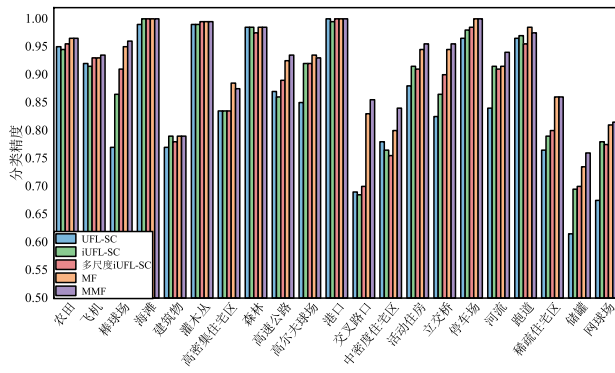


图7 UCM数据集中不同分类方法对于不同场景类别的分类精度

由图7可知,基于非监督特征学习的UFL-SC算法通过基于谱聚类的非监督学习提取特征,在海滩、灌木等场景达到了较高的分类精度,而本文提出的iUFL-SC则取得了更为有效的结果,尤其是在河流、储罐等场景类别优势明显。这表明本文算法采用不同大小的局部图像块自主“学习”到的两个参数包含的信息更为丰富全面,能在一定程度上解决场景内地物目标形状大小不同造成的难题。相比于前三种基于单特征的分类方法,基于多特征的MF方法和基于多特征多尺度的MMF在每一类的分类精度均大幅上升。这也说明了采用多尺度以及多特征融合的方法能够获取鉴别特征,实现更为全面的场景描述。

本文方法在UCM数据集上获得的混淆矩阵如图8所示。在混淆矩阵中,每行的数据表示同一场景样本被预测为各个类别标签的占比,其中空白位置表示为0,斜对角线的数据即为每类场景被正确预测的比例。由图8可得知,利用MMF算法可以获得比较高的分类精度,每类的分类精度大都在80%以上,在农田、飞机、海滩、灌木丛等类别中分类精度呈现较高水平达到100%。此外,对于UCM数据集中高密度住宅区,中密度住宅区和稀疏住宅区等场景高度相似、仅密集程度不同的类别,MMF算法亦取得不错的分类结果。

表2为UCM数据集上各种方法的分类精度对比。BoVW(SIFT)、BoVW(LBP)、UFL-SC是基于BoVW模型的单特征单尺度分类方法。SAL-pLSA和SAL-LDA融合了光谱、纹理和SIFT特征,并分别采用pLSA、LDA两种主题模型进行特征编码。文献[22]所提算法则采用串联的方式融合了SIFT、DisCLBP等四种特征。可发现,上述多特征融合算法的分类效果要好于单一属性特征的方法。MSDNN、M²LBP和MS-CLBP是基于多尺度的分类方法,相较单一尺度也取得更好的结果。本文的分类方法通过谱聚类非监督学习方法提取特征,并采用多特征与多尺度融合,因此明显提升了遥感图像场景分类性能。

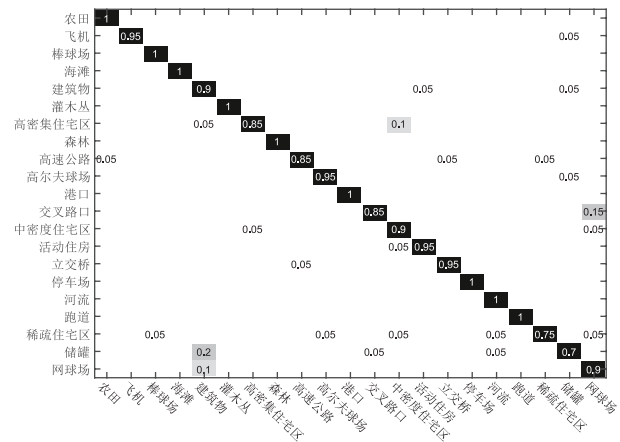


图8 MMF方法在UCM数据集上不同场景的分类混淆矩阵

表2 不同方法在UCM数据集上的分类精度对比

分类方法	分类精度(%)
BoVW ^[19]	71.86
BoVW(SIFT) ^[20]	74.12 ± 3.30
BoVW(LBP) ^[20]	77.12 ± 1.93
UFL-SC ^[16]	90.26 ± 1.51
SAL-pLSA ^[21]	87.62
SAL-LDA ^[21]	88.33
Multifeature concatenation ^[22]	89.48 ± 0.81
MSDNN ^[23]	91.34
M ² LBP ^[24]	91.34 ± 1.30
MS-CLBP ^[8]	90.60 ± 1.40
MMF(ours)	92.64 ± 0.96

4.5 WHU-RS19数据集实验结果与分析

WHU数据集中,五种算法在不同场景类别下的分类精度对比如图9所示。由图9可得知,在场景局部特征提取阶段采用多尺度的方式并形成联合局部特征能够明显提高分类效果,有效应对场景地物目标尺度多变的问题,例如居民区、高架桥等类别。此外,多特征融合的算法相比基于单一属性特征的分类方法,能够大幅提高最终的分类精度,在桥梁、足球场等类别具有明显优势。

WHU数据集中,基于MMF分类方法的混淆矩阵图10所示。不难发现,MMF算法可以获得比较高的分类效果,在草地、山区、公园等类别中分类精度呈现较高水平达到100%。其中,商业区、居民区和工业区均为房屋类建筑物,场景类间相似度高,极易造成错分的现象。通过采用MMF算法,得到的分类精度均超过75%,相比其他方法具有优势,说明本文方法对于复杂场景分类,具有更好的适应性,更有利实际应用。

表3为WHU数据集上的不同分类方法精度对比。相比于SIFT、CH等面向单一属性特征的分类方法,多特征级联^[22]、LRFF这类基于多特征融合的分类方法可

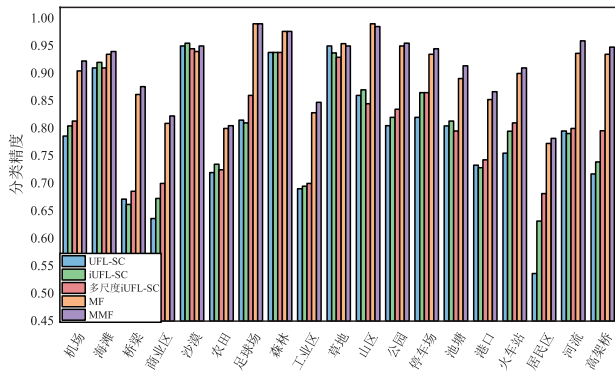


图9 WHU数据集中不同分类方法对于不同场景类别的分类精度

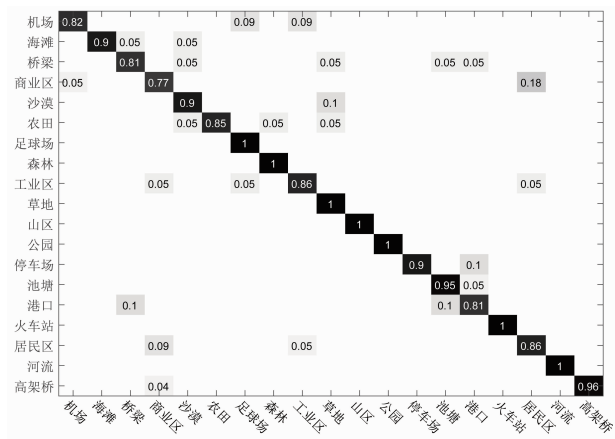


图10 MMF方法在WHU数据上的混淆矩阵

以获得更好的分类效果. 通过比较分类精度可知, 本文提出的 MMF 方法通过提取不同尺度的局部图像块并融合了不同属性的特征, 所以获得了更佳分类精度.

表 3 不同方法在 WHU 数据集上的分类精度对比

分类方法	分类精度 (%)
BoVW(SIFT) ^[20]	80.13 ± 2.01
BoVW(CH) ^[20]	66.55 ± 1.98
BoVW(LBP) ^[20]	74.97 ± 1.66
UFL-SC	79.40 ± 1.99
iUFL-SC	83.52 ± 1.73
multifeature concatenation ^[22]	90.79 ± 0.65
LRFF ^[25]	91.26 ± 0.47
MS-CLBP + BoVW ^[26]	89.29 ± 1.30
MMF(ours)	91.70 ± 0.96

5 结论

针对高分辨率遥感影像, 仅利用单一属性手工特征难以有效表达场景的问题, 本文提出了一种基于多尺度多特征的场景分类算法. 该算法首先设计一种改进的谱聚类非监督特征(iUFL-SC), 该方法通过自适应

的“学习”参数, 从而获取遥感场景的鉴别特征. 然后融合 iUFL-SC、LBP、SIFT 等三种多尺度局部特征并通过词袋模型进行编码. 该方法充分利用不同属性的特征, 从而实现更完备的场景描述. 在 UC Merced 和 WHU-RS19 高分辨率遥感图像场景数据集中的实验表明, 本算法相比于单一中层特征、多特征融合等分类方法有着更好的效果, 最高分类精度分别达到 92.64% 和 91.70%.

由于该算法需进行多尺度多特征提取及融合, 因此在局部特征提取及中层特征编码阶段将会耗费更多的时间. 此外, 本算法采用的中层编码方式为经典的 BoVW 编码模型, 若采用量化误差更小的编码模型能够提升效果, 因此下一步工作中将重点开展相关研究以进一步提升分类精度.

参考文献

[1] 何小飞, 邹峥嵘, 陶超, 等. 联合显著性和多层卷积神经网络的高分影像场景分类[J]. 测绘学报, 2016, 45(9): 1073 - 1080.
HE Xiao-fei, ZOU Zheng-rong, TAO Chao, et al. Combined saliency with multi-convolutional neural network for high resolution remote sensing scene classification [J]. Acta Geodaetica et Cartographica Sinica, 2016, 45(9): 1073 - 1080. (in Chinese)

[2] YANG Y, NEWSAM S. Bag-of-visual-words and spatial extensions for land-use classification[A]. Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems [C]. New York: ACM, 2010. 270 - 279.

[3] 杨赛, 赵春霞, 刘凡. 多核学习融合局部和全局特征的人脸识别算法[J]. 电子学报, 2016, 44(10): 2344 - 2350.
YANG Sai, ZHAO Chun-xia, LIU Fan. Fusion of local and global features using multiple kernel learning for face recognition[J]. Acta Electronica Sinica, 2016, 44(10): 2344 - 2350. (in Chinese)

[4] 张峰, 钟宝江. 基于兴趣目标的图像检索[J]. 电子学报, 2018, 46(8): 1915 - 1923.
ZHANG Feng, ZHONG Bao-jiang. Image retrieval based on interested objects[J]. Acta Electronica Sinica, 2018, 46(8): 1915 - 1923. (in Chinese)

[5] 钱晓亮, 李佳, 程臻, 等. 特征提取策略对高分辨率遥感图像场景分类性能影响的评估[J]. 遥感学报, 2018, 22(5): 758 - 776.
QIAN Xiao-liang, LI Jia, CHENG Gong, et al. Evaluation of the effect of feature extraction strategy on the performance of high-resolution remote sensing image scene classification[J]. Journal of Remote Sensing, 2018, 22(5): 758 - 776. (in Chinese)

- [6] HU J W, XIA G S, HU F, ZHANF L P. A comparative study of sampling analysis in the scene classification of optical high-spatial resolution remote sensing imagery [J]. *Remote Sensing*, 2015, 7(11): 14988 – 15013.
- [7] 胡运强, 曹云峰, 丁萌, 等. 基于分层搜索与局部约束线性编码的机场检测 [J]. *光学学报*, 2018, 38(8): 243 – 251.
HU Yun-qiang, CAO Yun-feng, DING Meng, et al. Airport detection based on a hierarchical architecture and locality-constrained linear coding [J]. *Acta Optica Sinica*, 2018, 38(8): 243 – 251. (in Chinese)
- [8] CHEN C, ZHANG B C, SU H J, et al. Land-use scene classification using multi-scale completed local binary patterns [J]. *Signal, Image and Video Processing*, 2016, 10(4): 745 – 752.
- [9] 柯圣财, 赵永威, 李弼程, 等. 基于卷积神经网络和监督核哈希的图像检索方法 [J]. *电子学报*, 2017, 45(1): 157 – 163.
KE Sheng-cai, ZHAO Yong-wei, LI Bi-cheng, et al. Image retrieval based on convolutional neural network and kernel-based supervised hashing [J]. *Acta Electronica Sinica*, 2017, 45(1): 157 – 163. (in Chinese)
- [10] DU B, XIONG W, WU J, et al. Stacked convolutional denoising auto-encoders for feature representation [J]. *IEEE Transactions on Cybernetics*, 2017, 47(4): 1017 – 1027.
- [11] XU Y H, ZHANG L P, DU B, et al. Spectral-spatial unified networks for hyperspectral image classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, 56(10): 5893 – 5909.
- [12] XU Y H, DU B, ZHANG F, et al. Hyperspectral image classification via a random patches network [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2018, 142: 344 – 357.
- [13] Penatti O A. B., Nogueira K, Santos J A D. Do deep features generalize from everyday objects to remote sensing and aerial scenes domains? [A]. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops* [C]. Boston: IEEE, 2015. 44 – 51.
- [14] Cheriadat A M. Unsupervised feature learning for aerial scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(1): 439 – 451.
- [15] ZHANG F, DU B, ZHANG L P, et al. Saliency-guided unsupervised feature learning for scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(4): 2175 – 2184.
- [16] HU F, XIA G S, WANG Z F, et al. Unsupervised feature learning via spectral clustering of multidimensional patches for remotely sensed scene classification [J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2015, 8(5): 2015 – 2030.
- [17] 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.
HE Fang, WANG Rong, YU Qiang, et al. Feature extraction of hyperspectral images of weighted spatial and spectral locality preserving projection [J]. *Optics and Precision Engineering*, 2017, 25(1): 263 – 273. (in Chinese)
- [18] 汪宇玲, 黎明, 李军华, 等. 基于 BoF 模型的多特征融合纹理图像分类 [J]. *北京航空航天大学学报*, 2018, 44(9): 1869 – 1877.
WANG Yu-ling, LI Ming, LI Jun-hua, et al. Texture image classification based on BoF model with multi-feature fusion [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2018, 44(9): 1869 – 1877. (in Chinese)
- [19] YANG Y, NEWSAM S. Spatial pyramid co-occurrence for image classification [A]. *Proceedings of the 2011 IEEE International Conference on Computer Vision* [C]. Barcelona: IEEE, 2011. 1465 – 1472.
- [20] XIA G S, HU J W, HU F, et al. AID: A benchmark data set for performance evaluation of aerial scene classification [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(7): 3965 – 3981.
- [21] ZHONG Y F, ZHU Q Q, ZHANG L P. Scene classification based on the multifeature fusion probabilistic topic model for high spatial resolution remote sensing imagery [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2015, 53(11): 6207 – 6222.
- [22] SHAO W, YANG W, XIA G S, et al. A hierarchical scheme of multiple feature fusion for high-resolution satellite scene categorization [A]. *Proceedings of the 9th International Conference on Computer Vision Systems* [C]. St Petersburg: Springer, 2013. 324 – 333.
- [23] 许凤晖, 慕晓冬, 赵鹏, 等. 利用多尺度特征与深度网络对遥感影像进行场景分类 [J]. *测绘学报*, 2016, 45(7): 834 – 840.
XU Su-hui, MU Xiao-dong, ZHAO Peng, et al. Scene classification of remote sensing image based on multi-scale feature and deep neural network [J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(7): 834 – 840. (in Chinese)
- [24] SHENG Y X, BIAN X Y. A novel spatial approach for classification of high-resolution image scene [A]. *Proceedings of the 12th World Congress on Intelligent Control and Automation* [C]. Guilin: IEEE, 2016. 1686 – 1691.
- [25] FERNANDO B, FROMONT E, MUSELET D, et al. Discriminative feature fusion for image classification [A]. *Proceedings of the 2012 IEEE Conference on Computer*

Vision and Pattern Recognition [C]. Providence: IEEE, 2012. 3434 – 3441.

- [26] HUANG L H, CHEN C, LI W, et al. Remote sensing image scene classification using multi-scale completed local binary patterns and fisher vectors [J]. Remote Sensing, 2016, 8(6): 483.

作者简介



黄 鸿(通信作者) 男,1980 年生于湖南新宁,2008 年获重庆大学博士学位,现为重庆大学教授、博士生导师. 主要研究方向为无人机遥感、遥感信息处理、流形学习、稀疏表示等.

E-mail: hhuang@cqu.edu.cn



徐科杰 男,1994 年生于浙江舟山. 2017 年于浙江师范大学获得学士学位,现为重庆大学博士研究生. 主要从事模式识别,图像处理,遥感影像分类等方面的研究.

E-mail: xukejie@cqu.edu.cn



石光耀 男,1988 年生于河南项城. 2015、2017 年于重庆大学分别获得学士和硕士学位,现为重庆大学博士研究生. 主要从事图像处理、遥感影像分类、机器视觉与目标追踪等方面的研究.

E-mail: shiguangyao@cqu.edu.cn