

徐培罡, 张海青, 王超, 等. 基于多重分割关联子的高分辨率遥感场景分类[J]. 地理科学, 2018, 38(2): 293-299. [Xu Peigang, Zhang Haiqing, Wang Chao et al. High Resolution Remote Sensing Image Classification Based on Multiple Segmentation Correlograms Model. Scientia Geographica Sinica, 2018, 38(2): 293-299.] doi: 10.13249/j.cnki.sgs.2018.02.016

基于多重分割关联子的高分辨率遥感场景分类

徐培罡¹, 张海青¹, 王超², 齐岗¹, 李杰¹, 吴静阳¹

(1. 国家测绘地理信息局第一航测遥感院, 陕西 西安 710054; 2. 内蒙古自治区基础地理信息中心, 内蒙古 呼和浩特 010010)

摘要: 高分辨率遥感影像提供了丰富的外观信息和空间结构信息, 广泛应用于土地利用分类当中, 源于文章领域的视觉词袋(Bag-of-Visual-Words, BoVW)模型现已成功应用于图像分类领域。传统的BoVW模型忽略了特征之间的空间布局信息和像素一致性信息, 提出多重分割关联子特征, 融合图像的外观信息、空间布局信息和像素一致性信息, 实验表明该方法能够获取优于许多经典的遥感图像特征的性能。

关键词: 多重分割; 空间信息; 像素一致性; 高分辨率遥感图像分类

中图分类号: P237 **文献标识码:** A **文章编号:** 1000-0690(2018)02-0293-07

随着遥感技术的发展, 各种传感器每天都在采集大量的遥感图像。高分辨率遥感图像(High-resolution Remote-sensing Images, HRIs)提供丰富的空间和纹理信息, 在各个领域中有着广泛的应用, 如军事^[1]、精准农业^[2]、采矿业^[3]、干旱监测^[4]、土地盐渍化^[5]、土地利用分类^[6]和灾害应急^[7]等, 越来越多的学者研究HRIs的信息自动提取和分类^[8-10]。

视觉词袋模型(Bag-of-Visual-Words, BoVW)是遥感图像场景分类中的经典方法, BoVW模型结合视觉词汇(Visual Word)和文本检索理论, 采用向量量化(Vector Quantization)的思想将图像的每个局部特征量化为视觉词典中的一个或多个视觉词汇, 并对所有局部特征的量化编码求和, 获取全局的向量表达形式, 即视觉词汇直方图。BoVW忽视了图像中局部特征的空间位置, 很多研究对其进行了改进, 如空间视觉词袋模型^[11]、空间金字塔(Spatial Pyramid Matching, SPM)^[12]、空间共现核(Spatial Co-occurrence Kernel, SCK)^[13]及SCK的金字塔扩展模型空间金字塔共现核(Spatial Pyramid Co-occurrence Kernel, SPCK)^[14]。这些特征采用规

则网格划分生成不同尺度的子区域, 统计并合并所有子区域的特征直方图。但是SPM及其扩展特征的空间信息对于目标或场景的方向较为敏感(如自然场景中天空位于地面上方), 而遥感图像则一般不具有上下或左右等方向性的位置关系, 而是以不同目标或地物的共现关系为主。视觉短语^[15]计算匹配特征的位置偏移, 从而探查相邻区域中视觉词汇的共现, 但空间约束过于严格, 适合于近似的图像, 而近似的遥感图像较少, 一般都是语义上相近。文献[16]基于图像的Correlogram矩阵, 提出一种空间信息的紧凑表达关联子模型, 联合了图像的外观和空间信息, 但模型忽略了遥感图像中的像素一致性, 而且关联子模型中的特征量化采用传统的K-均值聚类的方法, 存在较大的量化误差从而造成局部特征信息的损失。本文基于传统的关联子特征, 提出一种考虑像素一致性的关联子模型, 提高遥感图像场景分类的精度。

1 研究方法

1.1 基于多重分割Correlograms

本文基于传统Correlograms, 引入多重分割

收稿日期: 2017-02-14; 修订日期: 2017-05-07

基金项目: 国家自然科学基金项目(41271144)、2013年度河南省政府决策研究招标课题(2013B053)资助。[Foundation: National Natural Science Foundation of China (41271144), The 2013 Henan Provincial Government Decision-making Research (2013B053).]

作者简介: 徐培罡(1986-), 男, 陕西渭南人, 工程师, 主要从事摄影测量与遥感、数字城市及GIS应用开发与研究。E-mail: 370457225@qq.com

方法获取视觉单词之间的共现信息,提出一种基于多重分割 Correlograms 表达。多重分割 Correlograms 采用不同尺度大小的分割替代传统 Correlograms 中直接采用距离来设置距离核的方法,对于每张图像,通过 K 个不同尺度的分割获取视觉单词之间的空间共现信息和像素的一致性信息。多重分割 Correlograms 方法中多重分割基于 NCut (Normalized Cut) 的分割方法,NCut 是一种基于图分割的方法,在图像分割领域有着广泛的应用,能够保证同一分割块内最相似且不同分割块之间差别最大,方法的参数只需要指定分割个数。多重分割 Correlograms 方法通过选择 K 个不同的分割个数获取 K 个不同尺度的分割结果(图1)。NCut 方法采用区域内最相似而区域外相差最大的原则分割,不同尺度下的 NCut 能够得到不同大小、不同程度的像素一致性区域。如图1所示,不同尺度的分割基本上保留了飞机目标的完整性,而大尺度下飞机和周围具有相似性的地面会在同一区域,而相邻的飞机目标也会趋向于同一分割区域内(图1b~c),随着分割尺度的减小,单个飞机目标被分到单个区域内(图1d~e),而且在尺度更小的情况下目标的组成部分会被分割到不同的块内(图1f)。显而易见,外观上比较相似的块在空间上易

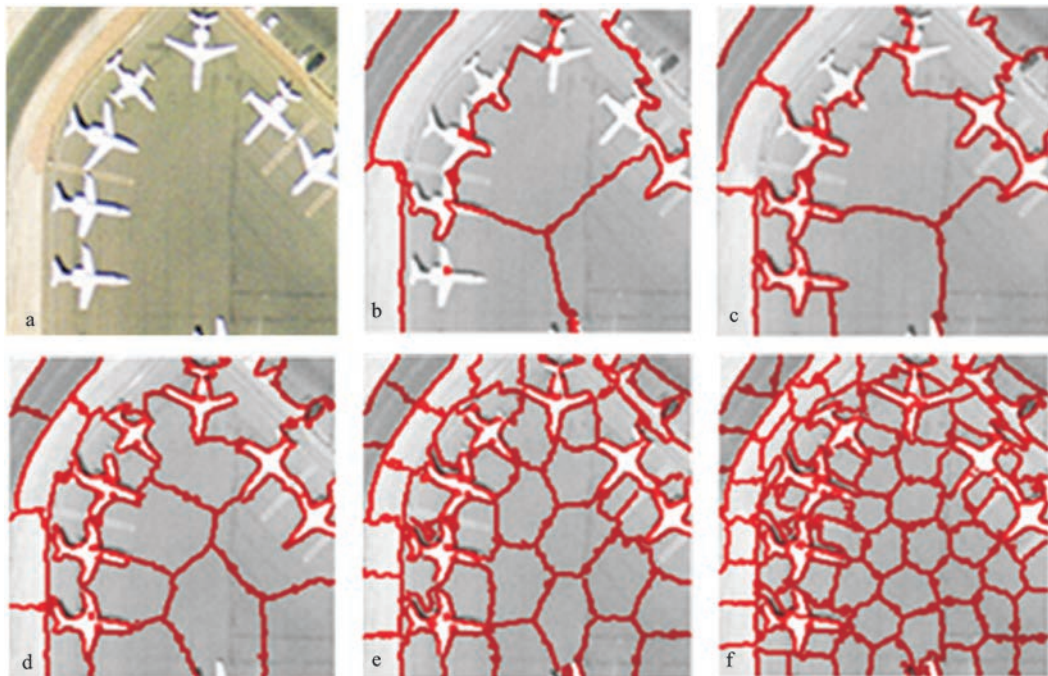
于相邻,表现为视觉单词对的共现,且从图中可以看到不同的分割大小能不捕捉不同尺度下的像素一致性。由此可见,多重分割能够捕获不同程度局部像素或特征的共现,而且随着分割尺度的减小,场景内相似区域或目标的像素空间上的共现信息越来越精细。

定义 Γ 为图像 I 为的分割, Γ_k 表示分割集合中分割结果 k (对应分割尺度 k ,如图1b~f),共 K 个不同的分割尺度。定义局部二值直方图 $B(\Gamma)$ 核实该为一个 $1 \times K$ 向量函数,用于获取同一分割块内共现的视觉单词。定义局部相关二值函数 $B(\Gamma_k, p, q)$,表示视觉单词 p 和 q 同时属于同一分割块,其中 Φ_i 表示视觉单词 i 的分割块标记。 $B(\Gamma_k, p, q)$ 的具体定义如下:

$$B(\Gamma_k, p, q) = \begin{cases} 1 & \text{if } \Phi_p = \Phi_q \\ 0 & \text{if not} \end{cases} \quad (1)$$

式中, Φ_i 为视觉单词 i 的分割块标记。

局部相关二值函数 $B(\Gamma_k, p, q)$ 对应原始 Correlograms 的局部相关直方图 $h(\Pi, p, q)$,局部相关二值函数采用不同尺度分割大小对应于局部相关直方图不同距离的核。由于 NCut 的分割块可能非常大,这样将影响视觉单词之间的相关性,因此多



a. 原始图像; b~f. 不同尺度(2^3 、 2^4 、 2^5 、 2^6 、 2^7)下的图像分割结果

图1 多重分割示例

Fig.1 Example of multiple scale segmentation

重分割 Correlograms 给出最大半径阈值,限定一定空间范围内的相关关系。我们定义最大半径阈值为 r ,则相关局部二值函数 $B(\Gamma_k, p, q)$ 的表达如式所示:

$$B(\Gamma_k, p, q) = \begin{cases} 1 & \text{if } \Phi_p = \Phi_q, \text{ dist}(p, q) < r \\ 0 & \text{if not} \end{cases} \quad (2)$$

式中, r 为最大半径阈值; $\text{dist}(p, q)$ 表示两个视觉特征之间的距离,本文定义其为不同的坐标维度下的最大差值,具体的定义如式下:

$$\text{dist}(p, q) = \max(|x_p - x_q|, |y_p - y_q|) \quad (3)$$

式中, x 和 y 对应局部视觉特征在图像中的位置。由于不同视觉单词在训练样本中出现个数的差异,我们采用同一标签下视觉单词的个数进行归一化,最终得到平均局部直方图 $L(\Gamma_r, i)$ 。平均局部直方图用于描述标签 i 下每个视觉单词在 k^{th} 分割块内的平均分布,具体的定义如下:

$$L(\Gamma_k, i) = \sum_{p \in \{p_i\}} \sum_{q \in \{q_j\}} \frac{B(\Gamma_k, p, q)}{|p_i|} \quad (4)$$

式中, $\{p_i\}$ 表示标签 i 下所有视觉单词的集合, $|p_i|$ 表示集合 $\{p_i\}$ 的基数; $\{q_j\}$ 表示标签 j 下所有视觉单词的集合, $|q_j|$ 表示集合 $\{q_j\}$ 的基数。

多重分割 Correlograms 最终得到 $T \times T \times K$ 的矩阵 $M(I)$ 。 $M(I)$ 是由 $L(\Gamma_r, i)$ 的 j^{th} 元素构成的,定义 $L(\Gamma_r, i, j)$ 为 $L(\Gamma_r, i)$ 的 j^{th} 元素,其中 $i = 1 \cdots T$, $j = 1 \cdots T$, $r = 1 \cdots K$ 。就像传统的 Correlogram 一样,多重分割 Correlogram 的元素是一个 $1 \times K$ 的向量,对应 $M(I)$ 的 i 行、 j 列,因此每个多重分割 Correlograms 拥有 $T \times T$ 个元素,每个多重分割 Correlograms 元素表达视觉单词对 p 和 q 随着分割尺度的变化,其空间相关性的变化信息。

1.2 基于多重分割的关联子模型

基于多重分割 Correlograms,构建一种融合外观、空间相关和像素一致性信息的模型,该模型在保持一定的判别能力的基础上,得到多重分割 Correlograms 一种紧凑表达模型,基于多重分割的关联子模型。这种紧凑表达不仅能够提高效率,而且有助于降低过拟合的风险。基于多重分割的关联子表达的高分辨率遥感图像分类模型具体流程如下:

1) 提取所有训练样本的特征并生成特征向量,采用经典的 K -means 聚类^[17]生成视觉词汇字典,查找训练样本中的所有描述子最近的视觉单

词做为特征的标签,最终统计所有视觉单词的频率,得到 BoVW 直方图表达。

2) 对所有训练样本进行多重分割,计算多重分割 Correlograms,收集训练样本中所有视觉单词对相关的多重分割 Correlograms 元素($1 \times K$ 向量)。

3) 采用 K -means 方法对所有的多重分割 Correlograms 元素进行聚类,最终得到的聚类中心就是对多重分割 Correlograms 元素的紧凑表达,即多重分割的关联子,从 Correlograms 到关联子的处理过程可以看作一种自适应的量化过程,选择少量的 Correlograms 元素样本近似地表示 Correlograms 矩阵,这种自动选择出的多重分割的关联子不仅捕获了空间相关性,而且保持了像素的一致性信息。

4) 所有训练样本的多重分割 Correlograms 元素映射至其相关的多重分割的关联子,统计所有关联子的频率,最终得到所有训练样本的多重分割的关联子直方图。

5) 合并所有训练样本的 BoVW 直方图和多重分割的关联子直方图,采用 SVM(Support Vector Machine,支持向量机)训练并生成分类器。

测试阶段基于训练完成的视觉字典和关联子字典,计算测试样本的 BoVW 表达、多重分割 Correlograms 矩阵和多重分割的关联子表达,测试图像具体的处理流程如下:

1) 提取图像的特征描述算子,采用视觉字典进行量化,计算 BoVW 直方图,主要用于捕获遥感图像的外观信息。

2) 对图像进行多重分割,采用量化过的特征向量,计算多重分割 Correlograms 矩阵。

3) 采用关联子字典对所有的多重分割 Correlograms 矩阵进行量化,计算多重分割的关联子统计直方图,直方图在考虑了像素一致性的前提下,描述视觉单词之间潜在的空间相关性。

4) 合并 BoVW 直方图和多重分割的关联子直方图,获取一种集成了外观、空间和像素一致性信息的遥感图像表达。本文没有考虑 BoVW 直方图和关联子直方图的权重信息,认为两者权重是相同的。

5) 采用训练阶段生成的 SVM 分类器对最终的基于多重分割的关联子模型进行分类,得到最终的分类决策。

2 实验结果与分析

2.1 特征和数据集

本文实验采用一种适合杂乱场景特性的稠密特征^[8],这种的稠密特征属于 SIFT 特征的稠密版本,采用稠密的规则网格替代传统 SIFT 方法中的关键点,计算每个格网块的 SIFT 特征。研究^[12]实验证明稠密 SIFT 特征在场景的检索和分类中表现优于传统的 SIFT 特征。本文的实验中,稠密 SIFT 特征采用 16×16 像素大小的格网,格网之间间隔 8 个像素,随机从训练样本中抽取一定数目的特征集合,并采用 K-means 聚类方法获取视觉字典。

为了验证算法的有效性,本文实验采用 UC Merced Land Use 数据集。UC Merced 数据集 (UCM) 来源于 USGS (United States Geological Survey) 的高分辨率航空正射图像^[9],一共 21 个类别,每个类别包含 100 张大小 256×256 像素、分辨率 0.3 m 的光学遥感图像。

2.2 实验相关设置

本文的实验对于两个数据集进行分割,其中一半训练样本,一半测试样本,采用测试样本的分类精度度量模型的有效性。实验采用 3 种经典模型作为参考,包括传统的 BoVW 模型、SPM 模型和传统的关联子模型,所有的算法均采用稠密 SIFT 特征。其中 SPM 算法采用 3 层金字塔,生成 3 种不同尺度的子区域;对于传统的关联子模型,我们采用核最大半径 80 和核数目 10;对于基于多重分割的关联子模型,我们设置分割尺度数目 $K=6$,其中每个分割尺度的大小 $N=\{4, 8, 16, 32, 64, 128\}$ (分别对应 2 的指数 $n=\{2, 3, 4, 5, 6, 7\}$)。

2.3 视汇表大小

词汇表的大小分类性能的影响是至关重要的,为了测试视觉词汇表大小和关联子词汇表大小对分类精度的影响,本文选择了不同大小的视觉词汇表和关联子词汇表,均以 50 为间隔从 50 到 300。

不同方法在不同大小视觉词汇下 UCM 数据集的分类精度如图 2 所示。从图中可以看出,随着视觉词汇表数量的增加,所有方法的分类精度均逐渐提升,其中传统的关联子模型(图 2 中的关联子模型)和基于多重分割的关联子模型在不同大小的视觉词汇表下均优于其他的方法。视觉词汇表较小时,SPM 与关联子和基于多重分割的关联子模型精度相当,且均优于 BoVW 的方法。随着

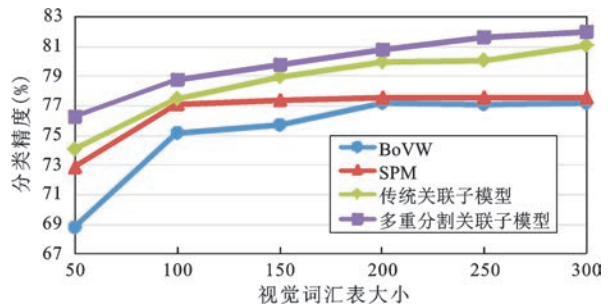


图2 不同大小视觉字典下的分类精度

Fig.2 Classification accuracy under different size of visual vocabulary

视觉词汇数量的增多,SPM 和 BoVW 模型之间的性能差异越来越小,而 BoVW 和 SPM 两个模型与关联子和多重分割的关联子两个模型的性能差异则越来越大。对于遥感图像场景来说,SPM 在视觉词汇表较小的情况下能够表现出不错的辨识力,但随着视觉词汇表的增大,SPM 中空间信息的作用慢慢减弱,尤其是在视觉词汇表大于 200 时,SPM 和 BoVW 模型分类精度基本相当。而关联子和基于多重分割的关联子模型在视觉词汇表大小达到 300 时,分类精度依然有提升,这说明 SPM 适合于视觉词汇表较小的情况,关联子和基于多重分割的关联子模型在不同视觉词汇大小下均能够表现出不错的性能。与传统的关联子模型相比,由于考虑了像素一致性,对于不同的视觉词汇大小,基于多重分割的关联子模型均能得到优于传统关联子模型的性能,从而证明了像素一致性信息能够提升关联子模型对于遥感图像的描述能力。总体上,相对于 BoVW、SPM 和传统的关联子模型,本文方法精度分别提升了 4.67%、3.21% 和 1.30%。

除了视觉字典的大小,本文提出的基于多重分割的关联子模型的精度还与关联子词汇表的大小有关。图 3 展示了 UCM 数据集视觉词汇关于关联子词汇表大小的函数变化关系。从图中可以看出,随着关联子词汇表大小的增加,不同的关联子词汇表大小下的模型分类精度快速提升,逐渐趋于最大值,最后开始慢慢降低,对于同一视觉词汇表大小,当分类性能达到最优时,说明模型的外观信息和(像素一致性约束下的)空间信息达到平衡。对于不同的词汇表大小来说,模型的最优关联子词汇表大小随着视觉词汇表大小增大而增大,这也说明了过大的关联子词汇表可能

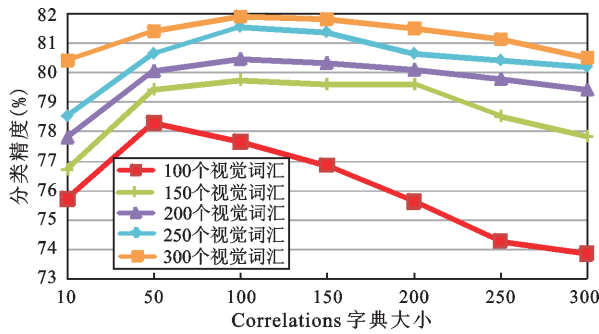


图3 不同视觉词汇表大小和不同关联子词汇表大小下的分类性能

Fig.3 Classification accuracy in different size of visual vocabulary and correlation vocabulary

造成更具有辨识力的外观信息损失。关联子直方图能够捕获局部特征之间的空间信息和一定的外观信息,外观信息对于高分辨率遥感图像的分类分辨来说相对更重要一些;关联子融合的空间信息描述局部视觉特征之间的共现关系,只是外观信息的辅助信息。因此,随着关联子词汇表大小的增加,外观信息相对于空间信息的比重也随之减小,从而降低了多重分割的关联子的描述力,尤其是关联子词汇表大小超过视觉词汇表大小的情况。对于UCM数据集,我们选择的最优视觉词汇表大小和关联子词汇表大小分别为300和100。

2.4 混淆矩阵

本文采用混淆矩阵描述基于多重分割的关联子模型对于UCM数据集的分类精度。如图4所示,UCM数据集下的农场、丛林、港口和停车场类别,本文方法能够获得超过98%的分类精度;而最易混淆的遥感图像场景类别如下:储存罐和建筑物、棒球场和高尔夫球场以及密集居民区和中密度居民区。

3 结论与讨论

基于多重分割的关联子模型针对传统关联子模型捕获空间信息忽略像素一致性信息的问题,从多重分割的方法出发,在传统关联子模型量化局部特征之间共现信息的基础上,考虑不同尺度下的像素一致性信息,从而得到一种更具辨识力的遥感图形特征表达。本文提出的多重分割关联子模型不仅能够融合外观信息、空间信息和像素一致性信息,而且方法采用K均值聚类对多重分割Correlograms进行量化并统计直方图,从而得到一种更加紧凑的遥感图像特征表达,不仅能够有效降低模型的计算量,而且降低了模型过拟合的风险,从而提高模型的分类的精度和效率。但是多重分割关联子中特征量化采用K均值方法可能局部特征量化误差,从而造成局部特征信息的丢失,我们将针对该问题做进一步的深入讨论和研究。

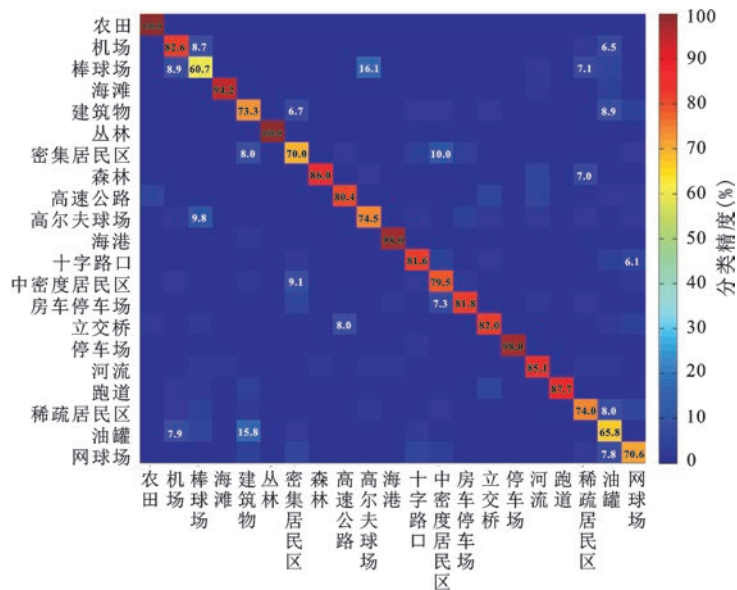


图4 基于多重分割的关联子模型最优参数下的混淆矩阵(单位:%)

Fig.4 Confusion matrix for the land-use dataset using multiple segmentation-based correlation model (unit: %)

参考文献(References):

- [1] Eismann M T, Stocker A D, Nasrabadi N M. Automated hyperspectral cueing for civilian search and rescue[J]. *Proceedings of IEEE*, 2009, 97(6): 1031-1055.
- [2] Goel P K, Prasher S O, Patel R M et al. Classification of hyperspectral data by decision trees and artificial neural networks to identify weed stress and nitrogen status of corn[J]. *Computers and Electronics in Agriculture*, 2003, 39(2): 67-93.
- [3] Iordache M D, Bioucas-Dias J M, Plaza A. Sparse unmixing of hyperspectral data[J]. *IEEE Transactions Geoscience and Remote Sensing*, 2011, 49(6): 2014-2039.
- [4] 周磊, 武建军, 张洁. 以遥感为基础的干旱监测方法研究进展[J]. *地理科学*, 2015, 35(5): 630-636. [Zhou Lei, Wu Jianjun, Zhang Jie. Remote sensing-based drought monitoring approach and research progress. *Scientia Geographica Sinica*, 2015, 35(5): 630-636.]
- [5] 张利华, 翟靖超, 李珊, 等. ALOS 融合影像质量评价及其土地盐渍化应用研究[J]. *地理科学*, 2015, 35(6): 798-804. [Zhang Lihua, Zhai Jingchao, Li Shan et al. Quality evaluation and land salinization classification application on ALOS image fusion. *Scientia Geographica Sinica*, 2015, 35(6): 798-804.]
- [6] 赵书河, 冯学智, 赵锐, 等. 资源 1 号卫星数据在国土资源调查中的应用评价[J]. *测绘通报*, 2001(6): 20-21. [Zhao Shuhe, Feng Xuezh, Zhao Rui et al. Study of applying CBERS-1 data to Investigation of land resources. *Bulletin of Surveying and Mapping*, 2001(6): 20-21.]
- [7] 李金香, 李亚芳, 李帅, 等. 面向地震应急准备的居民地遥感提取及量化分析[J]. *地理科学*, 2016, 36(11): 1743-1750. [Li Jinxiang, Li Yafang, Li Shuai et al. Remote sensing extraction and quantitative analysis of residential area for earthquake emergency preparedness. *Scientia Geographica Sinica*, 2016, 36(11): 1743-1750.]
- [8] 张倩, 黄昕, 张良培. 多尺度同质区域提取的高分辨率遥感影像分类研究[J]. *武汉大学学报信息科学版*, 2011, 36(1): 117-121. [Zhang Qian, Huang Xin, Zhang Liangpei. Multiscale image segmentation and classification with supervised ECHO of high spatial resolution remotely sensed Imagery. *Geomatics and Information Science of Wuhan University*, 2011, 36(1): 117-121.]
- [9] 沈忱, 祁昆仑, 刘文轩, 等. 基于 FSFDP-BoV 模型的遥感影像检索[J]. *地理与地理信息科学*, 2016, 32(1): 55-59. [Shen Chen, Qi Kunlun, Liu Wenxuan et al. Remote sensing image retrieval research based on FSFDP-BoV model. *Geography and Geo-Information Science*, 2016, 32(1): 55-59.]
- [10] 程诗尧, 梅天灿, 刘国英. 顾及结构特征的多层次马尔科夫随机场模型在影像分类中的应用[J]. *武汉大学学报(信息科学版)*, 2015, 40(9): 1180-1187. [Cheng Shiyao, Mei Tiancan, Liu Guoying. Application of multi-level MRF using structural feature to remote sensing image classification. *Geomatics and Information Science of Wuhan University*, 2015, 40(9): 1180-1187.]
- [11] Cao Y, Wang C, Li Z et al. Spatial-bag-of-features[C]// *Darrell T. Computer vision and pattern recognition (CVPR)*, 2010 IEEE conference on. Washington, D C, USA: IEEE Computer Society, IEEE, 2010: 3352-3359.
- [12] Lazebnik S, Schmid C, Ponce J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories[C]// *Fitzgibbon A, Taylor C J, Le Cun Y. Computer vision and pattern recognition*, 2006 IEEE computer society conference on. Washington, DC, USA: IEEE Computer Society, IEEE, 2006: 2169-2178.
- [13] Yang Y, Newsam S. Bag-of-visual-words and spatial extensions for land-use classification[C]// *Abbadi A E, Mokbel M. Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. New York, USA: ACM, 2010: 270-279.
- [14] Yang Y, Newsam S. Spatial pyramid co-occurrence for image classification[C]// *Metaxas. Computer vision (ICCV)*, 2011 IEEE international conference on. IEEE, 2011: 1465-1472.
- [15] Zhang Y, Jia Z, Chen T. Image retrieval with geometry-preserving visual phrases[C]// *Metaxas. Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE conference on. Washington, DC, USA: IEEE Computer Society, IEEE, 2011: 809-816.
- [16] Fei-Fei L, Perona P. A bayesian hierarchical model for learning natural scene categories[C]// *Schmid C. Computer vision and pattern recognition*, 2005. IEEE computer society conference on. Washington, DC, USA: IEEE Computer Society, 2005: 524-531.
- [17] 杨志民, 化祥雨, 叶娅芬, 等. 金融空间联系及 K-means 聚类中心等级识别研究——以长三角为例[J]. *地理科学*, 2015, 35(2): 144-150. [Yang Zhimin, Hua Xiangyu, Ye Yafen et al. Spatial combination of finance and center level identify based on K-means clustering: A case study of the Changjiang River Delta. *Scientia Geographica Sinica*, 2015, 35(2): 144-150.]

High Resolution Remote Sensing Image Classification Based on Multiple Segmentation Correlograms Model

Xu Peigang¹, Zhang Haiqing¹, Wang Chao², Qi Gang¹, Li Jie¹, Wu Jingyang¹

(1. *The First Institute of Photogrammetry and Remote Sensing, Xi'an 710054, Shaanxi, China;*

2. *Geometric Center of Inner Mongolia, Hohhot 010010, Inner Mongolia, China*)

Abstract: High resolution remote sensing images are increasingly applied in land use classification problems. However, it is a difficult task to recognize the semantic category for the complex background and multiple land-cover classes. The bag-of-visual-words model has been successful in scene classification, but ignores pixel homogeneity in land use remote sensing images. In this article, we present a multiple segmentation-based correlation feature to jointly integrate appearance, spatial correlation, and pixel homogeneity. We use a dense feature representation to detect spurious features resulting from clutter, which has been demonstrated that dense features work better for scene classification. These dense features are Scale-Invariant Feature Transform descriptors using a dense regular grid instead of interest points to extract features. The visual vocabulary is formed by K -means clustering of a random subset of patches from the training set. A multiple segmentation-based correlogram, which is a matrix express spatial co-occurrences of features, encoding both the local and global shape of visual words and robust with respect to basic geometric transformations and occlusions, were extracted. The corresponding correlogram elements for each pair of visual word labels in training images are collected and clustered using K -means. The set of cluster centers are multiple segmentation-based correlatons, which are a set of representative multiple segmentation-based correlogram elements. Therefore, multiple segmentation-based correlatons compress the co-occurrences information contained in a multiple segmentation-based correlogram without loss of discrimination accuracy. Finally, the concatenated histograms of images, which describe the underlying spatial correlation of visual words considering pixel homogeneity in the image region, are used as input feature vectors for the SVM classifier. The effectiveness of the multiple segmentation-based correlaton models was tested on a ground truth image dataset of 21 land use classes manually extracted from high-resolution remote-sensing images. The experimental results demonstrate that our improved correlaton model can promote classification and outperforms existing methods for the jointly integration of appearance, spatial correlation, and pixel homogeneity information.

Key words: multiple scale segmentation; spatial information; pixel homogeneity; scene classification based on high-resolution remote sensing