

基于不同特征的随机森林极化 SAR 图像分类^①



陈媛媛¹, 郑加柱², 魏浩翰¹, 张荣春², 欧翔¹

¹(南京林业大学 土木工程学院, 南京 210037)

²(南京邮电大学 地理与生物信息学院, 南京 210023)

通讯作者: 陈媛媛, E-mail: hhucyy@126.com

摘要: 近些年, 利用计算机对极化 SAR 图像进行分类逐渐成为遥感领域的一个研究热点. 本文采用全极化 SAR 数据, 利用不同的特征提取算法提取特征, 并基于随机森林模型最终实现对江苏沿海滩涂的分类. 首先采用 H/α 和 Freeman 两种分解算法提取极化特征参数, 采用灰度共生矩阵提取纹理特征参数, 然后将提取的所有特征进行不同的组合, 构成不同的特征集; 最后采用随机森林模型对不同特征集合进行分类和精度评估. 结果表明仅用纹理特征对沿海滩涂进行分类时效果较差; 利用极化分解提取出的散射特征进行分类的结果要优于矩阵元素特征的分类结果; 综合了极化散射特征和纹理特征的组合方式在沿海滩涂的分类中可以取得最优的分类结果, 总体精度和 Kappa 系数可以达到 94.44% 和 0.9305, 表明极化 SAR 图像中蕴含的不同方面的特征在分类中具有一定的互补性.
关键词: 极化 SAR; 极化分解; 特征提取; 随机森林; 分类

引用格式: 陈媛媛, 郑加柱, 魏浩翰, 张荣春, 欧翔. 基于不同特征的随机森林极化 SAR 图像分类. 计算机系统应用, 2019, 28(8): 183-189. <http://www.c-s-a.org.cn/1003-3254/7013.html>

Tidal Flat Classification Based on Random Forest Model Using Different Features of Polarimetric SAR

CHEN Yuan-Yuan¹, ZHENG Jia-Zhu², WEI Hao-Han¹, ZHANG Rong-Chun², OU Xiang¹

¹(College of Civil Engineering, Nanjing Forestry University, Nanjing 210037, China)

²(School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: The classification of polarimetric SAR images by computer has become a research hotspot in remote sensing. In this study, the fully polarimetric SAR data is used to extract characteristics by different algorithms, and the classification of tidal flat of Jiangsu coastal is realized. Firstly, the polarimetric scattering characteristics are extracted by H/α and Freeman decompositions, and the texture features are extracted by gray level co-occurrence matrix. Then, all the extracted features are combined to form different feature sets. Finally, the random forest model is used to classify and accurately evaluate with different feature sets. The study shows that using only texture features to classification achieves a poor performance. The classifications using the scattering features extracted by polarimetric decompositions are better than that of matrix element features. The combination of polarimetric scattering and texture characteristics can obtain best classification in coastal tidal flat, and the overall accuracy and Kappa coefficient are 94.44% and 0.9305, respectively. It indicates that the characteristics of different aspects contained in fully polarimetric SAR image have certain complementarity in the classification of coastal area.

Key words: polarimetric SAR; polarimetric decomposition; feature extraction; random forest; classification

① 基金项目: 江苏省测绘地理信息科研项目 (JSCHKY201708); 江苏省自然科学基金 (BK20180779); 南京林业大学青年科技创新基金 (CX2018015)

Foundation item: Surveying and Mapping Geographic Information Project of Jiangsu Province (JSCHKY201708); Natural Science Foundation of Jiangsu Province (BK20180779); Young Scientific and Technological Innovation Fund of Nanjing Forestry University (CX2018015)

收稿时间: 2019-01-30; 修改时间: 2019-02-27; 采用时间: 2019-03-06; csa 在线出版时间: 2019-08-08

遥感图像计算机分类一直是遥感领域的研究热点,分类精度的提高对于国土资源的监测与保护具有重要的意义^[1]. 近些年,随着遥感技术的飞速发展,各种传感器平台不断发射升空,为地表监测提供了丰富的数据来源^[2,3]. 极化 SAR 尤其是全极化 SAR 图像作为新的遥感手段,由于具有多个极化通道,可以获得更加丰富的地表信息,因此逐渐被用于地表分类及信息提取中^[4,5]. 目前,极化 SAR 图像散射信息提取主要采用极化分解的手段,不同的极化分解方法提取的特征参数对地物的敏感性也不尽相同^[6]. 极化 SAR 图像中除了散射信息,还包含有丰富的纹理信息^[7]. 目前的研究大多基于其中的某一类特征进行分类或信息提取,如果将极化 SAR 中的不同特征结合起来,势必会提高地表的分类精度. 因此,本文以江苏沿海滩涂为实验区域,不仅采用 H/α 和 Freeman 两种分解算法提取 SAR 图像中的极化特征参数,而且采用灰度共生矩阵提取纹理特征;然后将提取的所有特征进行不同的组合;最后采用随机森林模型对不同特征集合进行滩涂的分类和精度评估. 结果表明仅用纹理特征对沿海滩涂进行分类时效果较差;利用极化分解提取出的散射特征进行分类的结果要优于矩阵元素特征的分类结果;综合了极化散射特征和纹理特征的组合方式在沿海滩涂的分类中可以取得最优的分类结果,总体精度和 Kappa 系数可以达到 94.44% 和 0.9305,表明极化 SAR 图像中蕴含的不同方面的特征在滩涂分类中具有一定的互补性.

1 极化 SAR 图像特征提取

1.1 极化分解特征

本文采用经典的 H/α 分解^[8-10] 和 Freeman^[11,12] 分解来提取极化 SAR 图像中的极化散射特征参数. H/α 分解是对相干矩阵进行基于特征值/特征向量的分解,相干矩阵的形式如下:

$$T_3 = \sum_{i=1}^3 \lambda_i \cdot \mu_i \mu_i^{*T} \quad (1)$$

式中, λ_i 为相干矩阵的特征值,且 $\infty > \lambda_1 \geq \lambda_2 \geq \lambda_3 > 0$, μ_i 为与特征值对应的特征向量.

利用该分解可以得到以下参数:

$$H = - \sum_{i=1}^3 p_i \log_3 p_i \quad (2)$$

$$\alpha = \sum_{i=1}^3 p_i \alpha_i \quad (3)$$

$$A = \frac{\lambda_2 - \lambda_3}{\lambda_2 + \lambda_3} \quad (4)$$

$$\lambda = \lambda_1 + \lambda_2 + \lambda_3 \quad (5)$$

其中, $P_i = \lambda_i / \sum_{n=1}^3 \lambda_n$; λ_i 是相干矩阵的特征值, α_i 和 λ_i 分别表示各像素点的散射类型和各散射类型出现的概率. H 为散射熵,表示各散射机制在总散射过程中所占的比重,描述了散射过程的随机性; α 为散射角,代表平均散射机制类型; A 为各向异性度,描述 H/α 分解中两个较弱散射机制之间的关系^[13]. λ 为平均散射强度,包含了散射熵和散射角中没有的信息,这些信息可用于区分散射机制相似但散射强度不同的地物^[14].

Freeman3 分解是由 Freeman 和 Durden 提出的一种典型的基于非相干分解的分解算法^[11,12]. 它将协方差矩阵 C_3 分解为体散射、二面角散射和表面散射等三种散射机制的线性组合的形式,如下:

$$C_3 = f_s \langle [C_3] \rangle_{surface} + f_d \langle [C_3] \rangle_{double} + f_v \langle [C_3] \rangle_{vol} \quad (6)$$

其中, $\langle [C_3] \rangle_{surface}$ 对应表面散射; $\langle [C_3] \rangle_{double}$ 对应二面角散射; $\langle [C_3] \rangle_{vol}$ 体散射. f_s 、 f_d 、 f_v 分别对应于体散射分量的贡献、二面角散射分量的贡献及表面散射分量的贡献. 三种散射体对应的功率为:

$$P_s = f_s(1 + |\beta|^2) \quad (7)$$

$$P_d = f_d(1 + |\alpha|^2) \quad (8)$$

$$P_v = \frac{8}{3} f_v \quad (9)$$

1.2 纹理特征

灰度共生矩阵是像元距离和角度的矩阵函数,它通过计算图像中的一些距离和一定方向的两点灰度之间的相关性来反映图像在方向、间隔、变化幅度以及快慢上的综合信息^[7,15]. 本文根据灰度共生矩阵计算了四种统计量,分别是熵、差异性、均匀性、角二阶矩,公式如下:

$$\text{均匀性: } hom = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} \frac{g(i,j)}{1 + |i - j|} \quad (10)$$

$$\text{差异性: } dis = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} |i - j| \cdot g(i,j) \quad (11)$$

$$\text{熵: } en = - \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} g(i,j) \cdot \log(g(i,j)) \quad (12)$$

$$\text{熵角二阶矩: } ASM = \sum_{i=0}^{Ng-1} \sum_{j=0}^{Ng-1} g(i, j)^2 \quad (13)$$

2 随机森林模型

随机森林是近些年发展起来的一种机器学习模型^[16-18]. 该模型的理论基础是决策树, 是对决策树进行组合得到的, 即在变量和数据的使用上进行随机化, 生成很多决策树分类模型 $\{h(X, \theta_k), k=1, \dots\}$, 每棵树之间是没有关联的, 其中参数集 θ_k 为独立同分布的随机向量, 在自变量 X 给定时, 每个决策树分类模型都采用投票的方法产生最优的结果. 当原始数据进入随机森林后, 每棵决策树都对其进行分类, 最后取所有树中出现频率最高的分类结果作为最终结果.

① 采用自助法 (Bootstrap) 有放回地从原始训练数据集中随机抽取 k 个自助样本集, 利用这 k 个样本集构建 k 棵决策树. 在这一过程, 每次未被抽取的样本组成 k 个袋外数据 (Out-Of-Bag, OOB);

② 设有 N 个特征, 则在每一棵树的每个节点处随

机抽取 n 个特征 ($n \leq N$), 通过计算每个特征蕴含的信息量, 选择一个分类能力最强的特征进行分裂, 这样决策树的某一个叶子节点要么是无法继续分裂的, 要么里面的所有样本都指向同一个分类;

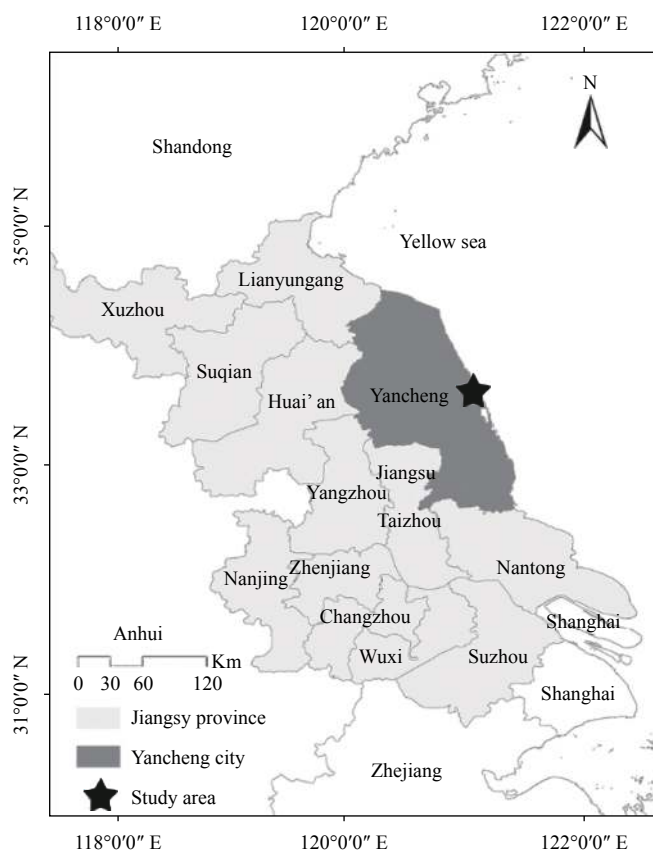
③ 每棵树都不进行剪枝, 使其最大限度地生长;

④ 所有决策树组成随机森林, 随机森林构建后, 将新的样本输入分类器中, 对于每个样本每棵决策树都对其类别进行投票, 分类结果按决策树投票数决定.

3 实验与分析

3.1 实验数据及实验方案

本实验选用 L 波段全极化 ALOS PALSAR 数据对江苏沿海滩涂进行分类, 研究区域如图 1 所示. 此外, 还选取了 2008 年 5 月获取的 QuickBird 高分光学影像及 Google Earth 卫星图像作为辅助数据, 以便对结果进行目视判读. 根据图像地物覆盖类型, 将研究区域分为水体、道路、鱼塘、沙滩、农田、芦苇和盐蒿等 6 种典型地物.



(a) 研究区域位置



(b) 研究区域对应的 Pauli 图像

图 1 研究区域: (a) 研究区域位置; (b) 研究区域对应的 Pauli 图像

通过 H/α 分解、Freeman3分解和灰度共生矩阵算法一共得到7个极化特征和4个纹理特征,加上相干矩阵的6个相干矩阵元素,一共得到17个特征,根据表1对它们进行不同的组合.然后采用随机森林模型对特征向量进行分类,本文实验的技术流程如图2所示,具体步骤为:(1)利用多视处理、滤波算法等对原始图像进行预处理;(2)对滤波后的图像进行 H/α 和Freeman分解,获取极化分解特征;(3)利用灰度共生矩阵计算Pauli图像的纹理特征;(4)特征组合;(5)选择训练样本和验证样本;(6)利用随机森林算法对表1中的不同特征组合向量进行分类;(7)计算分类精度.

表1 不同特征组合方式

特征集	特征参数
纹理特征	Hom 、 Dis 、 En 、 ASM
相干矩阵特征	$T11$ 、 $T12$ 、 $T13$ 、 $T22$ 、 $T23$ 、 $T33$
H/α 分解特征	H 、 α 、 A 、 λ
Freeman分解特征	$Freeman_Odd$ 、 $Freeman_Dbl$ 、 $Freeman_Vol$
所有特征	H 、 α 、 A 、 λ 、 $Freeman_Odd$ 、 $Freeman_Dbl$ 、 $Freeman_Vol$ 、 Hom 、 Dis 、 En 、 ASM 、 $T11$ 、 $T12$ 、 $T13$ 、 $T22$ 、 $T23$ 、 $T33$

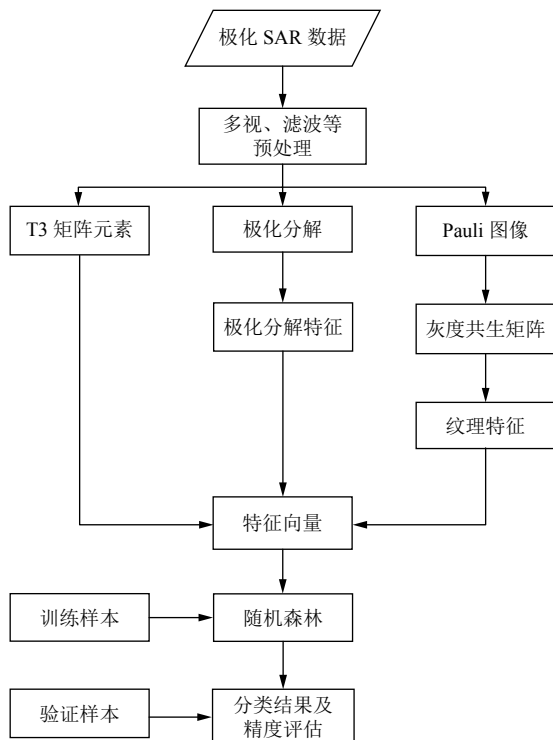


图2 技术路线

3.2 实验结果及精度评估

为了减少监督分类中选择样本对最后结果产生的影响,本次实验中选取同一组训练样本,选取随机森林模型对5组不同的特征组合实施分类,结果如图3所示.并利用同一组验证样本计算每个分类结果中的总体精度、生产者精度、用户精度和Kappa系数,从而进行定量评估(表2).

通过与高分辨率QuickBird影像和谷歌地图卫星影像比较进行目视判读.仅利用极化SAR图像蕴含的纹理特征进行分类时,所有地物被分成了海洋、道路和鱼塘三种,且总体精度和Kappa系数都很低,仅分别为30.38%和0.1508,而农田、沙滩、芦苇与盐蒿等根本无法识别出来.当把相干矩阵的6个元素作为极化特征向量组进行分类时,精度相较于仅利用纹理特征的分类结果总体精度提高了31.31%,Kappa系数提高了0.3629,且能把部分农田、沙滩、芦苇与盐蒿等识别出来,但是生产者精度和用户精度这两个指标在道路这一地类上却有大幅下降,大部分道路被误分成了农田和沙滩,说明这些地物在极化SAR图像中呈现较为相似的散射特性.通过图3(c)、图3(d)及表2可以看出,当对利用极化分解算法提取出的极化特征集进行分类时,总体精度与相干矩阵元素得到的分类结果相比提升了12%以上,Kappa系数提升了0.17以上,这说明了极化分解可以挖掘出极化SAR矩阵元素所不能表达的信息.而比较 H/α 分解和Freeman分解这两种分解算法的分类结果及精度可以看出,利用Freeman模型分解算法得到的极化散射特征比 H/α 分解得到的极化特征对沿海滩涂更加敏感,分类效果更好,精度更高.当把所有特征组合在一起进行分类时,虽然有少部分鱼塘被误分为了海洋,这可能是因为均发生奇次散射的缘故,但总体精度达到了94.44%,Kappa系数达到了0.9305,所有地类的生产者精度和用户精度与其他方法相比均有所提高,这说明,利用不同分解算法提取的极化特征参数对不同地物的敏感性不同,综合极化SAR图像中的纹理信息和极化散射信息进行分类时可以有效地提高分类精度,它们在滩涂分类中具有一定的互补性.

4 结论与展望

本文利用不同的极化分解算法和灰度共生矩阵分

别从极化 SAR 图像中提取出极化特征参数和纹理特征参数,并将这些特征组成不同的特征集,然后利用随机森林方法对江苏沿海滩涂的全极化 ALOS PALSAR 影像进行了分类.实验表明,纹理特征除了对道路较为敏感,对其他地类的分类效果都比极化特征差;采用极

化分解算法提取的极化特征对滩涂的分类效果要优于矩阵元素特征;综合了极化散射特征和纹理特征的组合方式在沿海滩涂的分类中可以取得较好的分类结果,表明极化 SAR 图像中蕴含的不同方面的特征在滩涂分类中具有一定的互补性.

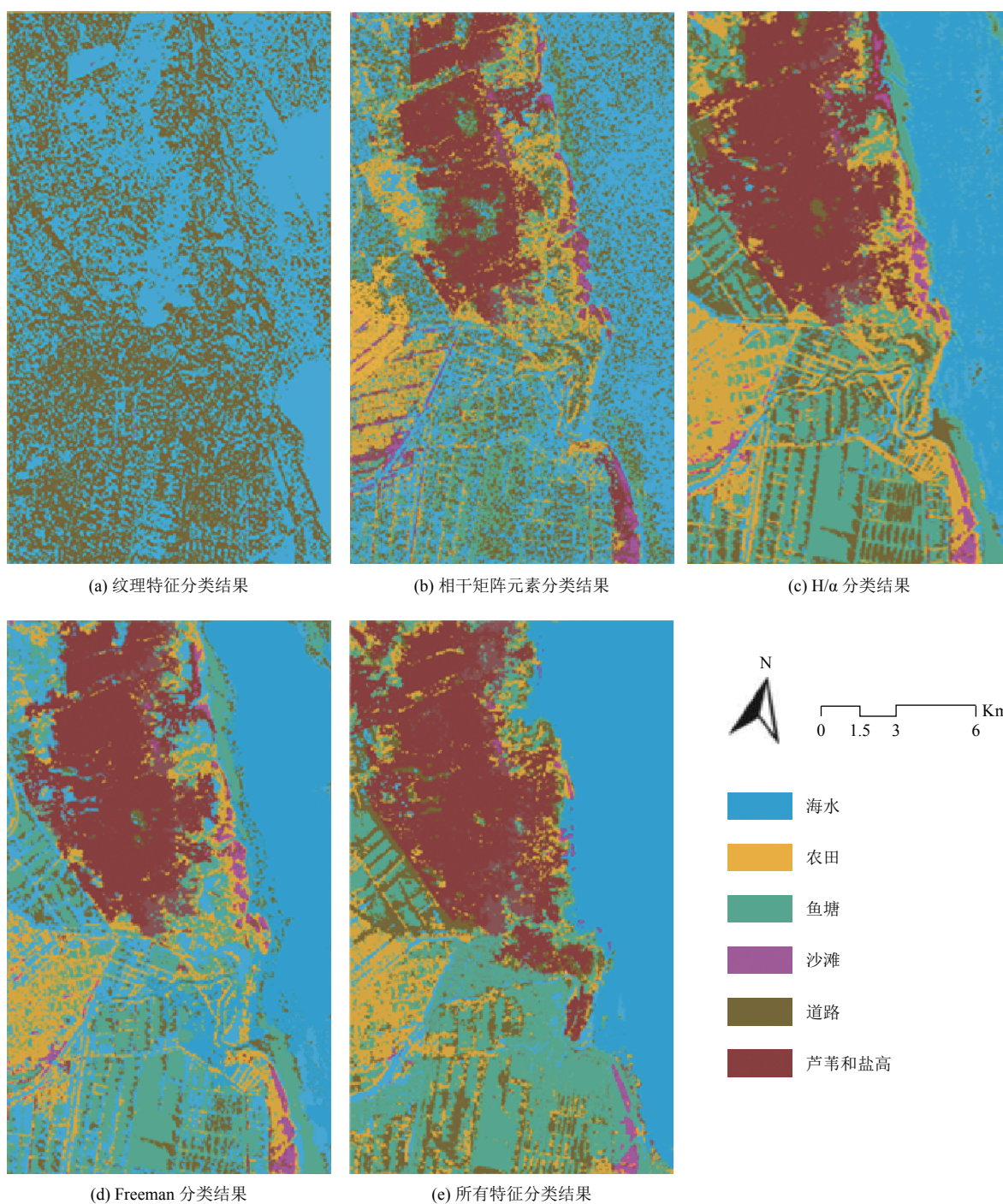


图3 分类结果

表2 不同特征组合下的分类精度

特征集	精度	地类					
		海水	道路	鱼塘	农田	沙滩	芦苇&盐蒿
纹理特征	生产者精度 (%)	82.85	51.79	4.03	0.00	0.00	0.00
	用户精度 (%)	31.48	25.66	55.56	0.00	0.00	0.00
	总体精度 (%)	30.38					
	Kappa 系数	0.1508					
相干矩阵元素	生产者精度 (%)	82.12	25.89	22.68	68.15	66.46	59.76
	用户精度 (%)	57.69	23.39	62.86	85.79	92.11	44.14
	总体精度 (%)	61.69					
	Kappa 系数	0.5137					
H/α	生产者精度 (%)	69.86	50.32	90.10	68.70	88.46	67.82
	用户精度 (%)	71.90	69.64	73.39	81.44	72.78	71.95
	总体精度 (%)	73.12					
	Kappa 系数	0.6868					
Freeman	生产者精度 (%)	60.95	79.46	94.76	100.00	79.11	60.98
	用户精度 (%)	97.09	45.88	85.77	82.91	100.00	56.18
	总体精度 (%)	78.5788					
	Kappa 系数	0.7369					
所有特征	生产者精度 (%)	95.26	94.64	98.39	98.97	86.71	89.02
	用户精度 (%)	99.62	89.08	94.94	94.12	97.16	81.11
	总体精度 (%)	94.44					
	Kappa 系数	0.9305					

参考文献

- 1 彭建, 王仰麟. 我国沿海滩涂的研究. 北京大学学报 (自然科学版), 2000, 36(6): 832-839. [doi: 10.3321/j.issn:0479-8023.2000.06.016]
- 2 詹雅婷, 朱叶飞, 苏一鸣, 等. 基于国土资源卫星的盐城海岸带生态环境变化调查. 国土资源遥感, 2017, 29(S1): 160-165. [doi: 10.6046/gtzyyg.2017.s1.27]
- 3 黄桂林, 张建军, 李玉祥. 辽河三角洲湿地分类及现状分析——辽河三角洲湿地资源及其生物多样性的遥感监测系列论文之一. 林业资源管理, 2000, (4): 51-56. [doi: 10.3969/j.issn.1002-6622.2000.04.013]
- 4 Chen YY, He XF, Wang J. Classification of coastal wetlands in eastern China using polarimetric SAR data. Arabian Journal of Geosciences, 2015, 8(12): 10203-10211. [doi: 10.1007/s12517-015-1940-2]
- 5 Buono A, Nunziata F, Migliaccio M, et al. Classification of the Yellow River delta area using fully polarimetric SAR measurements. International Journal of Remote Sensing, 2017, 38(23): 6714-6734. [doi: 10.1080/01431161.2017.1363437]
- 6 Chen YY, He XF, Wang J, et al. The influence of polarimetric parameters and an object-based approach on land cover classification in coastal wetlands. Remote Sensing, 2014, 6(12): 12575-12592. [doi: 10.3390/rs61212575]
- 7 邓滢, 张红, 王超, 等. 结合纹理与极化分解的面向对象极化 SAR 水体提取方法. 遥感技术与应用, 2016, 31(4): 714-723.
- 8 Lee J S, Pottier E. Polarimetric radar imaging: From basics to applications. Boca Raton: CRC Press, 2009.
- 9 Cloude S R, Pottier E. An entropy based classification scheme for land applications of polarimetric SAR. IEEE Transactions on Geoscience and Remote Sensing, 1997, 35(1): 68-78. [doi: 10.1109/36.551935]
- 10 Cloude S R, Papathanassiou K, Hajnsek I. An eigenvector method for the extraction of surface parameters in polarimetric SAR. Proceedings of CEOS SAR Workshop. Toulouse, France. 2000. 693-698.
- 11 Freeman A, Durden S L. Three-component scattering model to describe polarimetric SAR data. Proceedings of SPIE Conference on Radar Polarimetry. San Diego, CA, USA. 1992. 213-224.
- 12 Freeman A, Durden S L. A three-component scattering model for polarimetric SAR data. IEEE Transactions on Geoscience and Remote Sensing, 1998, 36(3): 963-973. [doi: 10.1109/36.673687]
- 13 Pottier E, Lee J S. Application of the 'H/A/alpha' polarimetric decomposition theorem for unsupervised

- classification of fully polarimetric SAR data based on the wishart distribution. Proceedings of SAR Workshop: CEOS Committee on Earth Observation Satellites. Toulouse, France. 2000. 335–340.
- 14 Fang C, Wen H. A new classification method based on Cloude Pottier eigenvalue/eigenvector decomposition. Proceedings of 2005 IEEE International Geoscience and Remote Sensing Symposium. Seoul, South Korea. 2005. 4.
- 15 高程程, 惠晓威. 基于灰度共生矩阵的纹理特征提取. 计算机系统应用, 2010, 19(6): 195–198. [doi: [10.3969/j.issn.1003-3254.2010.06.047](https://doi.org/10.3969/j.issn.1003-3254.2010.06.047)]
- 16 Liaw A, Wiener M. Classification and regression by random forest. R News, 2002, 2-3: 18–22.
- 17 詹国旗, 杨国东, 王凤艳, 等. 基于特征空间优化的随机森林算法在 GF-2 影像湿地分类中的研究. 地球信息科学学报, 2018, 20(10): 1520–1528. [doi: [10.12082/dqxxkx.2018.180119](https://doi.org/10.12082/dqxxkx.2018.180119)]
- 18 崔小芳, 刘正军. 基于随机森林分类方法和多源遥感数据的湿地植被精细分类. 测绘与空间地理信息, 2018, 41(8): 113–116. [doi: [10.3969/j.issn.1672-5867.2018.08.030](https://doi.org/10.3969/j.issn.1672-5867.2018.08.030)]

www.c-s-a.org.cn

www.c-s-a.org.cn