

# 相同地物多光谱遥感影像分类误差分析及其性能对比分析

杨金霖

成都理工大学数理学院, 四川 成都

收稿日期: 2023年8月14日; 录用日期: 2023年9月15日; 发布日期: 2023年9月22日

## 摘要

为了探究易混地物波谱的相近程度对分类的影响, 利用Landsat-8影像, 选取青海湖和洞庭湖作为研究区域。采用ROI剪裁, 选取同一地区(青海湖区)和跨区(青海湖区和洞庭湖区)遥感影像中的水, 阴影, 云, 植被, 陆地和建筑6种目标地物的三组样本集合, 以6种目标地物样本集合中的平均反射率作为各类目标地物的参照光谱曲线, 对比欧式距离, 曼哈顿距离, 余弦相似度三种不同衡量方式的分类结果精度, 实验结果表明: 1) 同一区域三种分类方式的分类精度中, 曼哈顿距离最好; 2) 不同区域波谱相近的地物中水的整体分类精度最高, 阴影和陆地较差。

## 关键词

多光谱遥感, 波谱分析, 欧式距离, 精度比较

# Analysis of Classification Error of Multispectral Remote Sensing Images of the Same Terrain and Its Comparative Analysis of Performance

Jinlin Yang

School of Mathematics and Science, Chengdu University of Technology, Chengdu Sichuan

Received: Aug. 14<sup>th</sup>, 2023; accepted: Sep. 15<sup>th</sup>, 2023; published: Sep. 22<sup>nd</sup>, 2023

## Abstract

In order to explore the influence of the similarity of the spectrum of easily mixable terrain on

文章引用: 杨金霖. 相同地物多光谱遥感影像分类误差分析及其性能对比分析[J]. 理论数学, 2023, 13(9): 2690-2696.

DOI: 10.12677/pm.2023.139276

classification, Qinghai Lake and Dongting Lake were selected as the study areas by using Landsat-8 images. Using ROI clipping, three sets of samples of water, shadow, cloud, vegetation, land and buildings in the same area (Qinghai Lake area) and cross-region (Qinghai Lake area and Dongting Lake area) remote sensing images were selected, and the average reflectivity of the six target feature sample sets was used as the reference spectral curve of various target features, and the accuracy of the classification results of three different measurement methods of Euclidean distance, Manhattan distance and cosine similarity was compared, and the experimental results showed that: 1) the classification accuracy of the three classification methods in the same area Manhattan is the best distance; 2) the overall classification accuracy of water in features with similar spectrum in different regions is the highest, and the shadow and land are poor.

## Keywords

Multispectral Remote Sensing, Spectral Analysis, Euclidean Distance, Accuracy Comparison

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



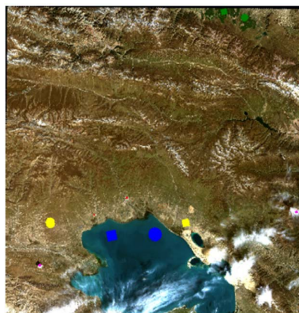
Open Access

## 1. 引言

多光谱遥感影像[1]记录了地物对电磁波的响应信息,由于不同地物反射率[2]不同,使得不同地物在影像中的特征存在差异,因此可以从包含多种地物遥感信息的影像中提取探测地物。目前遥感图像地物提取方法主要是依据地物的光谱特征差异构建,包括单波段阈值法、多波段谱间关系法和指数模型法等。在提取对象上,指数模型法主要应用于特定地物的专题信息提取,包括水体[3] [4]、居民点、植被[5]、耕地等典型地物,以及特殊区域中的典型地物,包括湿地[6]、海岸、冰雪等。然而该方法[7]在实际应用中对确定特征参数和光谱模型有一定局限性,因此只有与其他方法[8]进行互补融合提取,其识别精度和量化能力才会进一步提高。地物波谱特征[9]是遥感识别地物的重要依据。因此开展各种地物波谱特征测定和研究不仅是遥感的基础性工作,同时也是遥感应用研究的重要内容。

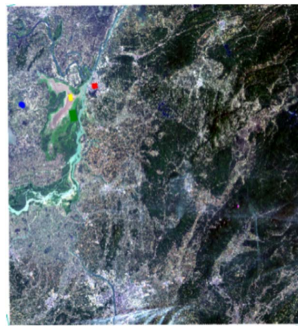
## 2. 数据获取及预处理

从地理空间数据云网站分别下载秋季青海湖影像和夏季洞庭湖 Landsat 8 遥感影像,如下图所示。



**Figure 1.** Remote sensing images of Qinghai Lake and their sample schematic diagram

**图 1.** 青海湖遥感影像及其样本示意图



**Figure 2.** Remote sensing images of Dongting Lake and their sample schematic diagram  
**图 2.** 洞庭湖遥感影像及其样本示意图

对多光谱遥感影像进行辐射定标和大气校正，得到反射率数据。对于图 1 青海湖的遥感影像，选取水，阴影，云，植被，陆地和建筑 6 种目标地物的样本，使用 ROI 剪裁出 10,000 个样本点作为样本 1 (以四边形标识) 的元素，6 种目标地物所有的样本点就构成了样本 1 的整体，其中水为蓝色，陆地为黄色，云为紫色，植被为绿色，阴影为黑色，建筑为红色。为了对比分析，在青海湖影像不同区域的相同目标地物选取第二个样本集合，记为样本 2 (以圆形标识)，样本 3 则选取洞庭湖遥感影像中 6 种目标地物。样本选取结果及其标识如图 1 和图 2 所示。

### 3. 实验结果与对比分析

#### 3.1. 分类方法简介

为了对比不同方法对目标地物的分类识别效果，采用余弦相似度、欧氏距离和曼哈顿距离三种方法对影像进行分类识别和目标提取。余弦相似度的度量方式[9]定义如下：

$$a \cdot b = \|a\| \cdot \|b\| \cos \theta,$$

给定两个属性向量， $A$  和  $B$ ，其余弦相似性  $\theta$  由点积和向量长度给出，定义如下所示：

$$\cos \theta = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}},$$

这里的  $A_i$ 、 $B_i$  分别代表向量  $A$  和  $B$  的各分量。相似性范围从 -1 到 1：-1 意味着两个向量指向的方向正好相反，1 表示它们的指向完全相同，0 通常表示它们之间是独立的。

假设两个  $n$  维数据点  $x = (x_1, x_2, x_3, \dots, x_n)$  和  $y = (y_1, y_2, y_3, \dots, y_n)$ ，欧式距离为：

$$\text{dist}(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} = \sqrt{(x - y)(x - y)^T}.$$

曼哈顿距离通常称为城市街区距离，两个  $n$  维向量  $x = (x_1, x_2, x_3, \dots, x_n)$  和  $y = (y_1, y_2, y_3, \dots, y_n)$  之间的曼哈顿距离定义为：

$$\text{dist}(x, y) = \sum_{i=1}^n |x_i - y_i|.$$

### 3.2. 结果分析

对于经过辐射定标和大气校正后的青海湖与洞庭湖遥感图像,采用 ROI 剪裁出水,阴影,云,植被,陆地,建筑 6 种目标地物构建样本,选择这 6 种地物反射率中的蓝,绿,红,近红外波段,短波近红外 1,短波近红外 2 共 6 个波段构建  $6 \times 6$  分类识别样本矩阵如下。其中  $A_i$  代表样本  $i$  的反射率平均值矩阵,  $a_{ij}$  代表第  $i$  类地物,第  $j$  个波段反射率值。

$$A_i = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{bmatrix}$$

为了对比相同地物在同一遥感影像不同区域以及不同遥感影像不同区域的分类识别效果,本研究选取同一地区(同一遥感影像)不同样本,即图 1 中的样本 1 和样本 2; 以及不同地区(分别为青海湖和洞庭湖对应的遥感影像)不同样本(即样本 1 和样本 3)的 6 种目标地物展开分类识别对比实验,并分析其性能,总结其中规律并指出存在的问题。

#### 3.2.1. 同一地区不同样本的分类识别结果分析

据图 1 所示,选取青海湖地区 6 种地物的两组不同样本集,并以样本集中像素反射率平均值构建 6 种目标地物的反射率曲线图,如下图 3 和图 4 所示。基于样本的反射率曲线,三种不同分类方法的分类识别效果如表 1 所示。

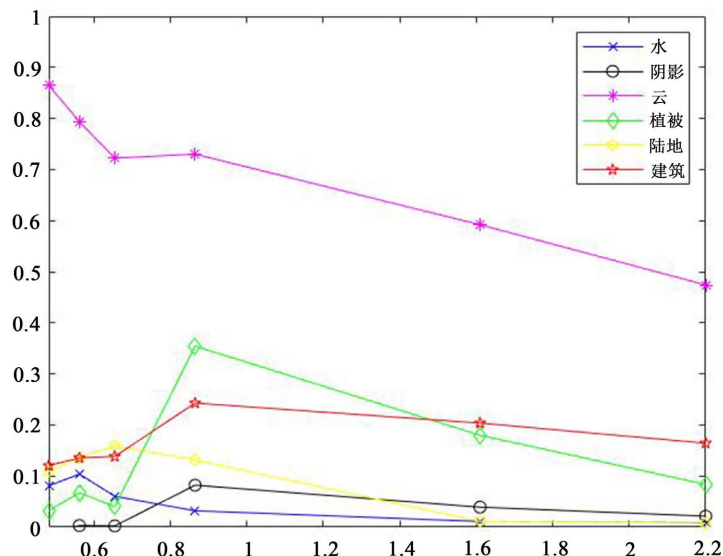


Figure 3. Spectral curves of six target features in Sample 1  
图 3. 样本 1 的 6 种目标地物的波谱曲线

实验 1 结果表明:对于同一地区不同区域的样本 1 和样本 2,以样本自身目标地物的反射率平均值来作为参照光谱曲线,余弦相似度,欧氏距离,曼哈顿距离都有很好的分类结果。其中曼哈顿距离的分类结果较优于余弦相似度和欧氏距离。

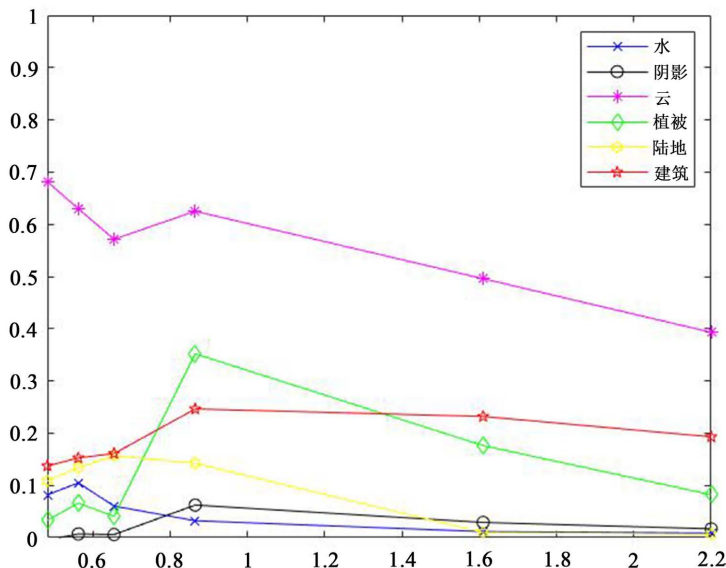


Figure 4. Spectral curves of six target features in Sample 2  
图 4. 样本 2 的 6 种目标地物的波谱曲线

Table 1. Classification accuracy of samples from different regions in the same region (unit: %)  
表 1. 同一地区不同区域样本的分类精度(单位: %)

地物种类	样本数	$A_1$ 测试样本 1 的正确率			$A_1$ 测试样本 2 的正确率			$A_2$ 测试样本 1 的正确率			$A_2$ 测试样本 2 的正确率		
		余弦相似度	欧氏距离	曼哈顿距离	余弦相似度	欧氏距离	曼哈顿距离	余弦相似度	欧式距离	曼哈顿距离	余弦相似度	欧氏距离	曼哈顿距离
水	10,000	100	100	100	100	100	100	100	100	100	100	100	100
阴影	700	94.5	97.3	98.4	97.8	96.3	89.4	98.7	92.7	94.2	100	97.1	97
云	2000	92.6	93.8	95.4	76.3	80	86.3	76.8	86.3	78.9	95.7	94.4	94%
植被	1000	72.6	86.6	73.6	64.8	74.6	69.8	35.7	73.6	71.5	57.3	74.6	78.8
陆地	10,000	99.4	99.8	99.8	97.9	97.4	96.3	79.7	98.1	98.0	97.5	99.1	99.6
建筑	500	85.8	87.6	88.6	86.8	86.8	87	80.5	82.8	81.8	83	83	82

### 3.2.2. 不同地区不同样本的分类识别结果分析

同一地区不同区域目标地物的反射率有微小扰动，因此导致表 1 的目标地物分类存在误差；例如以实验 1 样本 2 的阴影反射率平均值作为参照光谱曲线，对比以实验 1 样本 1 阴影本身的反射率为参照光谱曲线的分类结果，分类精度在三种分类方式上均存在微小误差。那么对于不同地区的目标地物来说，是否也存在目标地物反射率的扰动对分类结果有一定的误差？为了探究这种情况，进行实验 2。

据图 2 所示，选取洞庭湖地区 6 种地物的样本集，并以样本集中像素反射率平均值构建 6 种目标地物的反射率曲线图，如图 5 所示。基于样本的反射率曲线，三种不同分类方法的分类识别效果如表 2 所示。

从表 2 的实验结果来看，对于不同地区，以目标地物样本反射率的平均值作为参考波谱曲线，样本 1 的反射率平均值分类样本 3 与样本 3 的反射率平均值分类样本 1 在三种分类方法上分类结果均存在无法识别的情况。特别是表 2 中阴影和陆地的分类存在明显的识别不明。

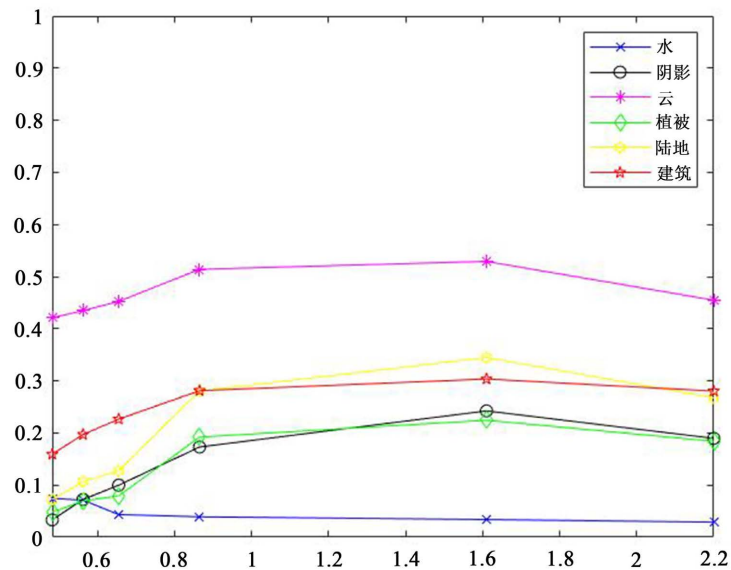


Figure 5. Spectral curves of six target features in Sample 3  
图 5. 样本 3 的 6 种目标地物的波谱曲线

Table 2. Classification accuracy of samples from different regions (unit: %)  
表 2. 不同地区样本的分类精度(单位: %)

地物种类	样本数	$A_1$ 测试样本 1 的正确率			$A_1$ 测试样本 3 的正确率			$A_3$ 测试样本 1 的正确率			$A_3$ 测试样本 3 的正确率		
		余弦相似度	欧氏距离	曼哈顿距离	余弦相似度	欧氏距离	曼哈顿距离	余弦相似度	欧氏距离	曼哈顿距离	余弦相似度	欧氏距离	曼哈顿距离
水	10,000	100	100	100	100	100	100	100	100	100	100	100	100
阴影	700	94.7	97.3	98.4	0	0	0	0	10	10	41	100	98
云	2000	92.6	93.8	95.4	29.3	82	96	38	96	70	94.6	56	83
植被	1000	72.6	86.6	73.6	100	74.6	68	42	100	56	100	28	100
陆地	10,000	99.4	99.8	99.8	0	0	0	0	0	0	99.8	99	100
建筑	500	85.8	87.6	88.6	38.6	62	82	85	97	61	92.6	99	96

#### 4. 结论

如何才能取得最优的反射率平均值使得对于所有的样本点判断精度最高是分类识别的关键问题之一。基于上述实验结果，得到以下结论。

- 1) 首先在获取采样样本时存在取点误差。原因是在制作样本标签时，可能存在对目标地物的标识错误，导致最终的分类判断错误。例如在上述实验 1 表 1 中植被的样本 1 和样本 2 本身的反射率平均值对自身的判别精度较低。对于该误差，可以通过多次取样使其降低甚至消除。
- 2) 数据精度误差。对于本文采用的三种判别方法，由于所取的目标地物反射率值在 0~1 之间，数据值较小，且余弦相似度和欧式距离都存在数值乘法运算，这就使得数据值更小，因此可能导致分类错误。加上程序实现时存在存储精度限制，也可能造成分类结果受数据精度误差影响。
- 3) 相较于实验 1，实验 2 中阴影和陆地的分类错误现象说明：即使是同种地物，不同地区存在的反射率微小差异，将对分类结果产生不可忽视的影响。如若简单以某个反射率为标准，分类结果可能会出



现重大错误。例如实验 2 中,以青海湖样本 1 阴影的反射率平均值作为参照光谱曲线验证洞庭湖样本 3 阴影的实验正确率为零和以洞庭湖样本 3 阴影的反射率平均值作为参照光谱曲线测试青海湖样本 1 阴影的实验正确率为零。针对这种现象,可以从指数法中得到启发,通过增大地物反射率之间的差距来突出某种地物使其达到想要的分类效果。

## 参考文献

- [1] 张明,黄双燕. 基于 Landsat-8 的遥感影像分类研究[J]. 测绘与空间地理信息, 2019, 42(1): 177-180.
- [2] 张亚梅. 地物反射波谱特征及高光谱成像遥感[J]. 光电技术应用, 2008, 23(5): 6-11+21.
- [3] Mishra, V.K. and Pant, T. (2022) Change Analysis of Water Area and Flood Mapping Using a Novel Water Index 2020 ( $WI_{2020}$ ) for Landsat Imagery. *Geocarto International*, **37**, 6391-6408.  
<https://doi.org/10.1080/10106049.2021.1939438>
- [4] Qi, Y.N., Dou, H.T. and Wang, Z.M. (2022) An Adaptive Threshold Selected Method from Remote Sensing Image Based on Water Index. *Journal of Physics: Conference Series*, **2228**, 1.  
<https://doi.org/10.1088/1742-6596/2228/1/012001>
- [5] 敖登, 杨佳慧, 丁维婷, 等. 54 种植被指数研究进展综述[J]. 安徽农业科学, 2023, 51(1): 13-21+28.
- [6] 文慧, 刘聚涛. 基于 Landsat 8 影像的地表水体指数提取方法研究——以鄱阳湖为例[J]. 江西水利科技, 2023, 49(4): 235-240+246.
- [7] 朱雨新, 李云梅, 张玉, 等. 基于遥感反射率的太湖优势藻识别方法[J]. 湖泊科学, 2023, 35(1): 73-87.
- [8] 胡荣明, 姚晓宙, 李鹏飞, 等. Landsat-8 影像的水体指数构建及稳定性研究[J]. 测绘科学, 2022, 47(5): 150-155.  
<https://doi.org/10.16251/j.cnki.1009-2307.2022.05.019>
- [9] 俞颖, 邵志荣, 林燕玲. 基于相似性度量的花卉图像检索方法研究[J]. 信息系统工程, 2019(6): 158-159.