

分类号：_____

密 级：_____

中国地质大学 硕士学位论文

面向对象遥感滑坡信息提取技术研究

学 号：120111304

硕 士 生：刘辰

学科专业：资源与环境遥感

指导教师：刘修国 教授

所在学院：信息工程学院

二〇一四年六月

学校代码: 10491

研究生学号: 120111304

中国地质大学 硕士学位论文

面向对象遥感滑坡信息提取技术研究

硕士生: 刘辰

学科专业: 资源与环境遥感

指导教师: 刘修国 教授

二〇一四年六月

A Dissertation Submitted to China University
of Geosciences for the Master Degree of Remote Sensing for
Natural Resources and Environment

Research on Object-Oriented Remote Sensing Landslide
Information Extraction Technology

Master Candidate: Liu Chen

Major: Remote Sensing for Natural Resources and
Environment

Supervisor: Liu Xiuguo Professor

China University of Geosciences


Wuhan 430074 P. R. China

中国地质大学（武汉）研究生学位论文原创性声明

本人郑重声明：本人所呈交的硕士学位论文《面向对象遥感滑坡信息提取技术研究》，是本人在导师的指导下，在中国地质大学（武汉）攻读硕士学位期间独立进行研究工作所取得的成果。论文中除已注明部分外不包含他人已发表或撰写过的研究成果，对论文的完成提供过帮助的有关人员已在文中说明并致以谢意。

本人所呈交的硕士学位论文没有违反学术道德和学术规范，没有侵权行为，并愿意承担由此而产生的法律责任和法律后果。

学位论文作者签名：



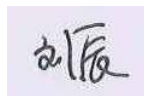
日 期： 年 月 日

中国地质大学（武汉）学位论文使用授权书

本人授权中国地质大学（武汉）可采用影印、缩印、数字化或其它复制手段保存本学位论文；学校可向国家有关部门或机构送交本学位论文的电子版全文，编入有关数据库进行检索、下载及文献传递服务；同意在校园网内提供全文浏览和下载服务。

涉密论文解密后适用于本授权书。

学位论文作者签名：_____



日 期：_____年 月 日

作者简介

刘辰，男，汉族，1988年8月出生，湖北汉阳人。2007年9月至2011年6月，就读于中国地质大学（武汉）信息工程学院，获遥感科学与技术学士学位。2011年9月保送中国地质大学（武汉）信息工程学院，攻读资源与环境遥感专业学术型硕士学位，主要研究方向为面向对象遥感信息提取。

在攻读硕士学位期间，完成了专业所规定的课程及学分，共修29个学分，其中学位课18个学分，选修课11个学分，平均成绩86.1分。主要完成了硕士专业英语、摄影测量原理、Visual C++程序设计、图像处理与分析、矩阵理论、地质灾害分析与防治、岩体稳定性评价与利用、岩土工程GIS技术、地质灾害风险分析与预警预报等16门课程的学习。

攻读硕士学位期间参加的科研项目包括：

1、2011.09~2012.08，参与《高分辨率对地观测重大专项：“高分地质灾害调查与监测示范原型系统开发”》项目（国家级，项目编号04-Y30B01-9001-12/15）

2、2012.09~2013.05，参与《高分辨率对地观测重大专项：“高分综合交通遥感应用示范系统先期攻关”》项目（国家级，项目编号07-Y30A05-9001-12/13）

3、2013.06~2014.04，参与《基于国产卫星的地质灾害调查监测系统研制与示范应用》项目（省部级，发改办高计[2012]2083号）

研究生期间发表论文情况：

1、刘辰，刘修国，陈启浩等. 面向对象滑坡信息提取中DEM空间分辨率影响分析[J]. 遥感技术与应用, 2014, 29(4). （待刊）

△2、Qihao Chen, Xiuguo Liu, **Chen Liu**, et al. Impact Analysis of Different Spatial Resolution DEM on Object-Oriented Landslide Extraction from High Resolution Remote Sensing Images. 9th International Conference on Natural Computation, Shenyang, 2013. （IEEE会议，EI Compendex 和 ISTP 双收录）

△3、刘修国, **刘辰**, 陈启浩等. DEM 空间分辨率对高分影像面向对象滑坡信息提取的影响分析. 第一届高分辨率对地观测学术研讨会论文集, 北京, 2012.

△4、陈启浩, 高伟, **刘辰**等. 基于高分辨率影像的面向对象滑坡信息提取——以“湖南省资兴市”为例. 第十届全国数学地质与地学信息学术研讨会论文集, 武汉, 2011.

面向对象遥感滑坡信息提取技术研究

硕士生：刘 辰

指导教师：刘修国

摘 要

滑坡是一种极具破坏的突发性地质灾害，是当今世界上造成经济损失和人员伤亡最为严重的自然灾害之一。国务院在 2006 年 2 月 9 日发布的《国家中长期科学和技术发展规划纲要（2006-2020 年）》中提出了国家高分辨率对地观测系统重大科技专项（简称高分专项）。遥感对地观测具有获取速度快、覆盖范围广、周期性等特点，是滑坡地质灾害调查研究的重要途径。本文依托“高分专项”中的“高分地质灾害调查与监测示范原型系统开发”、“高分综合交通遥感应用示范系统先期攻关”以及“基于国产卫星的地质灾害调查监测系统研制与示范应用”项目，围绕地质灾害调查与监测主体业务，针对面向对象遥感滑坡信息提取技术中如何设置分割参数、如何选择信息提取特征以及 DEM 空间分辨率对面向对象滑坡信息提取的影响等问题，取得了以下几个方面的研究成果：

1) 基于 SSC (Scale-Shape-Compactness) 的滑坡对象构建方法。提出一种 SSC 多尺度分割参数优化方法，通过对湖南省资兴市滑坡灾害区进行 SSC 分割参数优化实验，得到尺度 (Scale)、形状因子 (Shape) 和紧致度 (Compactness) 优化分割参数 Seg_{opt} ，将优化分割参数 Seg_{opt} 与默认分割参数 Seg_{def} 进行对比验证。通过实验结果得出，基于 SSC 的滑坡对象构建方法，在不具备先验知识的前提下，可以计算出一个分割参数，使得分割后产生的滑坡对象与实际目标轮廓最为接近。该方法对于面向对象遥感滑坡信息提取中如何设置多尺度分割参数具有重要的指导意义和应用价值。

2) 随机森林滑坡特征选择方法。提出一种基于对象的随机森林滑坡特征选择方法，通过 R 语言编写程序，对印度北部的北阿肯德邦地区滑坡信息提取过程中的地物进行随机森林特征选择，并计算得到所选取特征的贡献度。通过实验分析发现，利用随机森林特征选择方法得到的特征与根据专家经验选出的特征有着很强的一致性，且专家经验特征在随机森林选出的特征集中有着很高的贡献度。该方法在面向对象遥感滑坡信息提取中，对于寻找地物潜在诊断性特征具有重要的参考价值和启发意义。

3) 面向对象滑坡信息提取中 DEM 空间分辨率影响分析。为分析 DEM 分辨率对滑坡提取的影响并确定满足应用的分辨率要求，将原始 10m 分辨率 DEM 重采样为空间分辨率不同的 15m、20m、30m、50m、75m、100m，采用 DEM 及其衍生数据辅助高分遥感影像的面向对象滑坡信息提取方法进行实验。通过实验分析得出，当 DEM 分辨率小于 30m 时，对面积大于 5000m² 的滑坡能得到较好的识别和分类结果；大于 30m 时虽难以区分出滑坡类型，但通过调整部分规则参数仍能实现对滑坡的探测。该研究对于准确提取滑坡信息所

需 DEM 产品的选择具有重要的指导意义和参考价值。

本文的创新之处有：

1) 提出了一种基于 SSC (Scale-Shape-Compactness) 的滑坡对象构建方法。该方法针对面向对象滑坡信息提取多尺度分割中分割参数设置问题, 通过结合影像对象内均质度、对象间异质度以及对象面积创建的分割质量评价指标, 分别从多尺度分割的尺度参数、形状因子参数和紧致度参数三方面进行优化, 以得到面向对象遥感滑坡信息的最优多尺度分割参数。

2) 提出了一种基于随机森林的面向对象滑坡特征选择方法。该方法针对面向对象遥感滑坡信息提取中特征选择问题, 结合随机森林在特征选择方面的优势, 通过 R 语言编写的程序, 分别对遥感滑坡信息提取中各类地物进行随机森林特征选择运算, 得到各个特征的贡献度, 以指导面向对象滑坡信息提取时目标地物的诊断性特征的选取工作。

关键词：滑坡遥感；面向对象；多尺度分割；特征选择；数字高程模型（DEM）

Research on Object-Oriented Remote Sensing Landslide Information Extraction Technology

Master Candidate: Liu Chen

Supervisor: Liu Xiuguo

ABSTRACT

Landslide is a highly destructive sudden geological disaster, is one of the extremely serious natural disasters caused economic losses and casualties in the world. The State Council put forward a National High-Resolution Earth Observation System Major Science and Technology Projects (referred to High-Resolution Projects) in the "Long-Term Scientific and Technological Development Plan (2006-2020)" February 9, 2006. Earth observation remote sensing have access to fast, wide coverage, periodicity, and other characteristics, is an important way of investigation of landslides. This article relies on the "High-Resolution Survey and Geological Disaster Monitoring Demonstration Prototype System Development", "High-resolution Integrated Transport System in Advance Technological Demonstration of Remote Sensing Applications" and "Geological Disaster Monitoring System Development and Demonstration Application Based on Domestic Satellite" projects in the "High-resolution Projects", arounding the main business of investigation and monitoring of geological disasters, aiming at the problem of how to set the parameters, how to select the information extraction features and the impact of DEM spatial resolution on object-oriented landslide extraction, have the following aspects of the research findings:

- 1) Landslide Object Construction Method based on SSC (Scale-Shape-Compactness). Proposed a SSC multiscale segmentation parameter optimization method, by the SSC segmentation parameters optimization to the landslide disaster area in Hunan Zixing, get the optimization segmentation parameters (Seg_{opt}) in the field of Scale, Shape and Compactness, and then comparing the Seg_{opt} to the Seg_{def} (default segmentation parameters). By comparing the experimental results we can conclude that, in the premise of not have prior knowledge, we can calculate a segmentation parameters which makes the landslide object closest to the actual target size by the landslide object construction method based on SSC. This method has important

guiding significance and application value for how to set the multiscale segmentation parameters in object-oriented remote sensing landslide information extraction.

2) Random Forest Landslide Feature Selection Methods. Proposed an object-oriented random forest landslide feature selection method, using the application programmed by R language, selecting the features of classes in the landslide information extraction process on Uttarakhand region in northern India by random rorest feature selection methods, and then calculate the characteristics of the selected contribution. Experimental analysis found that, there is a strong consistency between the features selected by random rorest feature selection methods and features proposed by expert, and the expert features has a high contribution in the group of random forest selected features. This method has important reference value and instructive for seeking potential diagnostic features in object-oriented landslide information extraction.

3) Impact of DEM Spatial Resolution on Landslide Extraction using Object-Oriented Methods. In order to analyze the impact of DEM resolution on landslide extraction and to determine the resolution to the meet application requirements, we resample the 10m resolution DEM into 15m, 20m, 30m, 75m and 100m, and then adopt the object-oriented method combined with high-resolution remote sensing images and DEMs to extract the landslides information. The experimental results show that, when the DEM resolution is greater than 30 meters, we can obtain better recognition and classification results for the landslide with area greater than 5000m². When the resolution is less than 30 meters, it is difficult to distinguish between landslide types, but by adjusting parameter values we can still achieve the detection of landslides. This research has certain guiding significance and reference value for the selection of DEM spatial resolution on the landslides information extraction.

Here is the innovation of this article:

1) Proposed a Landslide Object Construction Method based on SSC (Scale-Shape-Compactness). This method aimed at the problem of how to set the multiscale segmentation parameters in Object-Oriented Remote Sensing Landslide Information Extraction, by combining the segmentation quality evaluation index which composed by homogeneity within objects, the heterogeneity between objects and the objects area, optimized in the Scale, Shape and Compactness in multiscale segmentation, and then obtain the optimal multiscale segmentation parameters in object-oriented remote sensing landslide information extraction.

2) Proposed an Object-Oriented Landslide Feature Selection Method based on Random Forests. This method aimed at the problem of how to select the information extraction features in Object-Oriented Remote Sensing Landslide Information Extraction, by combining the superiority of random forests in feature selection, using the application programmed by R language, selecting the features of classes in the remote sensing landslide information extraction process by random rorest feature selection methods, obtain the contribution of each feature, and then guiding the selection of diagnostic features.

Keywords: Landslide Remote Sensing; Object-Oriented; Multiscale Segmentation; Feature Selection; Digital Elevation Model (DEM)

目 录

第一章 绪论	1
§1.1 研究背景及意义.....	1
§1.2 国内外研究现状.....	1
1.2.1 遥感滑坡信息提取现状.....	2
1.2.2 最优分割参数选择现状.....	3
1.2.3 随机森林特征选择现状.....	4
§1.3 研究内容与技术路线.....	5
1.3.1 研究内容.....	5
1.3.2 技术路线.....	5
§1.4 论文章节内容安排.....	7
第二章 面向对象滑坡信息提取方法	8
§2.1 面向对象影像分析技术.....	8
2.1.1 影像分割.....	8
2.1.2 特征空间构建.....	10
2.1.3 特征选择.....	13
§2.2 遥感影像滑坡信息提取流程.....	15
§2.3 本章小结.....	16
第三章 基于 SSC 的滑坡对象构建方法	17
§3.1 多尺度分割的基本原理.....	17
§3.2 SSC 多尺度分割参数优化方法.....	19
3.2.1 分割质量评价.....	19
3.2.2 SSC 分割参数优化方法.....	21
§3.3 基于 SSC 的滑坡对象构建实验.....	21
3.3.1 实验区概况及数据介绍.....	21
3.3.2 基于 SSC 的滑坡对象构建.....	22
3.3.3 分割结果验证.....	27
§3.4 本章小结.....	27
第四章 随机森林滑坡特征选择方法	29
§4.1 随机森林的基本原理.....	29
§4.2 基于对象的随机森林滑坡特征选择.....	30
§4.3 随机森林滑坡特征选择实验.....	32
4.3.1 实验数据及预处理.....	32
4.3.2 选择候选影像对象特征.....	33

4.3.3 特征选择实验流程	35
4.3.4 实验结果与分析	35
§4.4 本章小结	44
第五章 面向对象滑坡信息提取中 DEM 空间分辨率影响分析	45
§5.1 研究区及数据描述	45
§5.2 面向对象的滑坡信息提取方法	45
§5.3 实验结果及分析	49
5.3.1 滑坡信息提取结果	49
5.3.2 精度评价	50
5.3.3 结果分析及结论	51
§5.4 本章小结	52
第六章 结论与展望	53
§6.1 结论	53
§6.2 展望	54
致 谢	55
参考文献	56

第一章 绪论

§1.1 研究背景及意义

国务院在 2006 年 2 月 9 日发布《国家中长期科学和技术发展规划纲要(2006-2020 年)》。纲要列出了中国在 2006 至 2020 年期间将要加快发展的 16 个重大专项,国家高分辨率对地观测系统重大科技专项(简称高分专项)位列其中。高分专项的主要实施目标和内容是:发展基于卫星、飞机和平流层飞艇的高分辨率先进观测系统;形成时空协调、全天候、全天时的对地观测系统;建立对地观测数据中心等地面支撑和运行系统,提高我国空间数据自给率,形成空间信息产业链。

在高分专项中,资源环境是四大应用领域之一。本研究选题来源于高分专项中的“高分地质灾害调查与监测示范原型系统开发”、“高分综合交通遥感应用示范系统先期攻关”以及“基于国产卫星的地质灾害调查监测系统研制与示范应用”等项目,主要围绕地质灾害调查与监测主体业务与“国家高分辨率对地观测系统”紧密结合,开展面向对象遥感滑坡信息提取技术研究。针对遥感信息提取技术中的如何设置分割参数、如何选择信息提取特征以及 DEM 空间分辨率对面向对象滑坡信息提取的影响等问题进行研究。

滑坡是一种极具破坏性的突发性地质灾害,是当今世界上除地震以外造成巨大经济损失和人员伤亡的自然灾害之一^[1]。遥感对地观测具有获取速度快、覆盖范围广、周期性强等特点^[2],使其成为滑坡地质灾害调查研究的重要途径之一。传统的滑坡遥感调查采用人工目视解译的方法,通过解译标志从遥感影像上获取滑坡信息^[3],该类方法不仅工作量大而且费时费力,不能很好的适用于突发性灾害的应急响应。随着高分辨率遥感卫星的快速发展,遥感影像展现出了较之以往更加丰富的几何结构和纹理信息。面向对象影像分析方法充分利用了影像的光谱、空间、纹理和邻域信息^[4],成为具有广阔前景的滑坡遥感信息提取方法。

§1.2 国内外研究现状

国内外现阶段对于面向对象滑坡信息提取主要有三个方面的发展趋势:首先,越来越多的学者将遥感技术应用于滑坡信息提取;其次,在高分遥感影像进行多尺度分割的过程中,为了使分割对象能够更好的表达滑坡边界,应该如何确定最佳滑坡对象分割参数;最后,影像分割产生了光谱、空间、纹理和邻域特征等数百个影像特征,由于特征维数过高,全部用于分类显然是不切实际的,那么应该如何从这些特征中选择对于滑坡信息提取贡献度最大的几个特征。根据上述提及的三个发展趋势,将当前面向对象滑坡信息提取的国内外研究现状归纳如下:

1.2.1 遥感滑坡信息提取现状

在科技发展日新月异的今天,越来越多的测量手段被用于地质灾害调查^{[5][6]}。早期的滑坡检测主要依靠高分辨率的航空影像,通过目视解译的方式来进行人工判读,然后再转绘到专题地图上^[7]。随着高分辨率卫星遥感影像在资源环境监测、灾害管理等领域广泛应用,高分影像信息提取方法成为近年来研究的热点,出现了利用高分辨率卫星影像替代早期的航摄影像进行滑坡检测^[8]。起初提取滑坡信息是利用影像分类、影像融合和变化检测等基于像素的遥感影像处理方法,但随着面向对象分析方法的提出,这种能整合对象多特征的分析方法被用于滑坡信息提取。

航空影像的目视解译和野外调查仍然是准备滑坡编录与制图的主要资料。虽然航空影像能够很精确的描述滑坡的细节,但它们通常不能及时地处理大范围的滑坡多发地区。与航空影像不同,卫星影像能实现大范围滑坡影响区域的损失评估,并且能评价滑坡发生地区的环境。最初,由于空间分辨率低的制约,早期的研究重点主要是大型滑坡探测。然而在最近的研究中,人们越来越多的使用高分辨率影像(如,QuickBird, IKONOS, WorldView-1, Cartosat-1/2, SPOT-5, ALOS-PRISM 等)来进行滑坡制图^[9]。王志华等在 2007 年以精确配准的滑坡前后遥感影像、数字地形和 DEM 共同作为解译基础,以人机交互方式解译和分析岩门村滑坡特征^[10]。赵祥等在 2008 年针对汶川地震灾区内滑坡泥石流多发的现象,分析了基于多源多时相高分辨光学遥感数据和 SAR 数据的滑坡泥石流提取方法^[11]。邓辉等在 2005 年利用且与全色波段融合后的 Quickbird-2 卫星多光谱影像作为遥感解译底图。从遥感影像上提取了白依庵滑坡的各要素,进行了稳定性评价和危害性分析,并提出了防治措施建议^[12]。

人工目视解译的方法提取滑坡信息,不仅工作量大而且费时费力。于是人们开始尝试用自动方法提取滑坡信息,比较常用的自动滑坡提取的方法是影像分类变化检测和影像融合。叶润青等 2007 年利用遥感图像处理软件 ENVI 对灾害点遥感影像分类处理及分析,探索滑坡灾害的光谱识别与特征提取。得出图像分类方法通过感兴趣区选择可以去除一些与滑坡灾害解译无关信息,一定程度上凸显滑坡解译有用信息,并进行灾害相关信息统计,有助于滑坡识别与特征提取^[13]。Borghuis 等 2007 年用非监督分类探测到了人工解译 63% 的滑坡^[14]。Nichol 等 2005 年用变化检测成功的将滑坡从光谱特征相近的裸露岩石和土壤中区分出来^[15]。然而,上面提到的自动方法是基于像素的,这些方法不能很好的表现滑坡的地质过程。因此,得到的结果表现为椒盐状,并且大多数不能实地证实。这些方法仅仅依赖影像光谱特征,而这又不是滑坡所特有的特征。除了光谱特征之外,滑坡的诊断特征还包括植被指数、坡度、坡向、流域、张力裂缝以及周围的人工地物。于是在之后的研究中,人们试图量化一些滑坡的诊断性特征。1988 年 Pike 通过一系列地形变量从 DEM 中计算几何特征,从而将滑坡与其周围地物分离^[16]。通过类似的方法,Iwahashi 和 Pike 在 2007 年利用 DEM 产生的坡度、地形纹理和当地的凸度进行地形的自动分类^[17]。姚鑫等 2008 年以碧罗雪山一处高山峡谷地区为例,采用 10m 空间分辨率的 SPOT-5 多光谱影像和 1:5 万地形图生成的 DEM 作为数据源,采用决策树分类方法进行了浅层崩滑体自动识别效果的检验^[18]。

面向对象的影像分类是一种知识驱动的方法,它通过专家知识结合滑坡的光谱、形态学和上下文等诊断性特征能精确地提取滑坡信息^[19]。因为滑坡发生在复杂的地质环境下,

将滑坡作为一个对象从周围环境中圈定出来显得尤为重要。影像分割是面向对象分析前必须进行的一个步骤,它将光谱性质相近的像素合并成一个对象^[20]。影像对象能表象更多滑坡诊断性的特征,相对于基于像素的方法,面向对象的方法在滑坡自动探测方面更具有潜力。然而,它们没有结合 DEM 数据进行分析,遥感数据结合 DEM 的衍生产品比单独的使用它们将得到更好的结果,Barlow 等人应用面向对象的方法,基于 Landsat ETM+影像在 Cascade 山地区开展了滑坡检测,在 DEM 数据的辅助下进行了滑坡遥感识别与制图^[19]。但由于影像分辨率低,得出的结果并不理想。Moine 等 2009 年通过面向对象分析方法利用形状、光谱、纹理和邻接特征从航空和卫星影像上探测滑坡,遗憾的是并没有考虑形态学参数^[21]。利用 DEM 数据辅助滑坡信息提取的研究还有,Iwahashi 和 Pike 于 2007 年利用 DEM 产生的坡度、地形纹理和当地的凸度进行地形的自动分类^[22]。胡德勇等 2008 年采用基于对象的高分辨率遥感图像滑坡检测方法,对马来西亚 Cameron 高原热带雨林地区的滑坡进行了提取^[23]。Martha 等 2010 年在印度 Mandakini 地区开展了基于多源数据(包括多光谱影像、DEM 等)的面向对象滑坡信息提取实验^[24]。沈永林等 2010 年将高分辨率航空影像和机载 LiDAR 数据生成的 DEM 相结合,开展针对海地地震诱发的滑坡识别研究工作^[25]。

以上研究清楚的表明面向对象分析方法在滑坡信息提取方面的潜能还未完全利用。结合光谱数据、数字高程模型及其衍生产品等多源数据的面向对象滑坡信息提取技术为快速准确的地质灾害分析与评价提供了可能性。DEM 在滑坡信息提取中起到了重要作用,可在实际应用中往往难以获取到研究区的高分辨率 DEM 数据,分析 DEM 分辨率对滑坡提取的影响具有重要意义。

1.2.2 最优分割参数选择现状

目前,影像分割已成遥感专题信息提取中不可或缺的技术,如何根据专题应用尺度将相邻的像元组成有意义的影像对象并融入对象的背景信息是影像分析成功的关键^{[26][27]}。考虑到多源遥感影像的复杂性,为获取最佳的影像对象,在进行影像分割之前,必需充分考虑以下几个要求^[28]: 1) 分割过程应该生成高度同质的分割区域,分割后的小区域具有最优的可分离性与代表性; 2) 几乎所有的影像对象属性如色调、纹理、形状等与邻域关联度或多或少与尺度有一定的依赖关系,类似尺度的影像空间结构在特征上有可比性; 3) 由于影像分析问题与给定尺度遥感数据的空间结构有关系,那么分割后影像对象的平均大小必须与感兴趣的尺度大小相适宜; 4) 分割过程应该具有普遍性,能适用于多种不同类型的数据与问题; 分割成果应该具有再生性等。

遥感影像是由空间结构上相互联结的像元组成的对象,反映了地表物体固有尺度。要理解影像对象如何在不同的尺度域之间相互作用,就需要一种技术既能自动生成遥感影像的影像对象,又能将这些影像对象按等级结构联接起来。这种技术就是多源影像的多尺度分割技术。

最近,多尺度分割技术已由 eCognition 公司开发并实现出来了^{[29][30]},这是一种自下而上的区域合并技术,由单个像素开始,持续进行合并步骤,直到完成用户定义的同质性准则。该技术利用模糊规则集理论根据感兴趣的尺度大小将影像分割为细尺度和粗尺度,更多关于该方法细节的解释参考文献^[31]。由于该方法能在不同尺度上产生高质量的影像对象,因此它已经应用于许多新的领域,如:车辆自动检测、变化检测研究、从扫描的纸质地图中提取信息以及地貌过程理解等^{[32][33][34][35][36][37]}。

如果影像基元对应于真实场景中的可见特征,那么通过处理方法或特征知识的有效结合来进行分割分类是唯一可能的。例如:Martha 等人证明了如何通过类似于坡度、地表曲率和长宽比等地貌参数,区分不同的滑坡类型^[24]。然而,这种性质必须适用于涉及到的所有滑坡。但是,产生的分割对象面临一个巨大的挑战,因为它依赖于—组用户定义的分割参数。这些值受限于所允许的内部影像异质性和对象的大小。这种方法效果很好,因为它所关注的特征在很大程度上是同质的且大小相近。然而,当特征从形状、大小和光谱表现方面来说是高度可变的话,那么初始分割通常是减少尝试和错误的方法。Möller 等人提出了一个比较工具来评估影像分割的准确性,并确定了用来改进土地利用和土地覆盖分类可靠性的最优参数^[38]。然而,根据 Zhang Hui 等人的研究,分割质量的客观评价是最好的选择,因为相关对象是不需要的,而且这种评价标准依赖于应用并且难以界定^[39]。最近,Dragut 等人已经发展了一种利用影像对象异质性局部标准差来估计尺度参数的工具^[40]。Espindola 等人提出一种目标函数来评估分割质量,并确定最优分割参数^[41]。目标函数计算对象内同质性和对象间异质性,因此它是一个很好的分割对象评价指标,尤其针对基于区域分割算法得到的对象^[39]。另外,在选择滑坡分割参数方面,Espindola 等人提出的方法^[41]相对于 Möller 等人的方法^[38]而言更加合适些,因为它不需要地形的先验知识。然而,Blaschke 等人^[42]强调了确定“有意义对象”的难度,他们是这样说的,“影像分割产生的有意义对象不可能适用于所有尺度以及任何现象”。

以上的研究表明在遥感影像分割中,为了使分割对象能够更好的表达目标地物边界,往往需要确定最优的分割参数。因此,如何选择最优分割参数对于,准确进行面向对象遥感滑坡信息提取具有重要的意义。

1.2.3 随机森林特征选择现状

随机森林(Random Forests)算法,是由 Breiman 于 2001 年提出,国内外公认效果较好的特征选择算法之一^[43]。随机森林是一个相对较新的非参数机器学习算法,可以处理多样化数据类型并且广泛采用于各种各样的应用^{[44][45]}。正如它名称所暗示的,随机森林利用多个决策树,即个别树的预测是整个森林的平均。

Sesnie 等人于 2008 年利用随机森林算法产生的重要性测量鉴定预测变量和生态学类别之间的关系,同时它也作为一个剔除对整体分类精度贡献低的预测变量的方法^[46]。Ghimire 等人于 2010 年结合 G 系数统计量、局部空间相关性和随机森林分类器组成空间依赖性,并将它用于基于像素的分类,研究结果表明,一个使用不同窗口大小来计算 G 系数统计的空间随机森林,在利用 Landsat ETM+影像区分土地覆盖类型时,表现为非空间随机森林^[47]。Prasad 等人于 2006 年将随机森林用于一系列潜在的未来气候情景使用多个生态变量来创建树种的预测图(Prasad *et al.* 2006),研究发现,在采取多种的评价方式的基础上,随机森林分类器比回归树分析、引导聚集树和多元自适应回归样条表现得略微要好一些^[48]。

相比于其它特征选择算法,这种基于集成分类算法结合由几个独立训练分类器(如决策树 DT)作出的预测,比使用单一分类器得到的分类精度要更加准确^[49]。随机森林算法能够处理相对于观测数目(n)而言更大数目的变量(v)。这些所谓 n 小 v 大的数据集,也被别人称为“维数灾难”^[50],常常涉及面向对象影像分析的应用,尤其是是在变量(对象特征)数目与考虑到的训练样本数目相当或超出的情况下。这样的性能特点使得随机森林特征选择成为面向对象影像分析应用的热门选择。

§1.3 研究内容与技术路线

1.3.1 研究内容

(1) 滑坡对象最优分割参数的研究

遥感影像分割主要由对象内同质性和对象间异质性来决定, 由于无法获取每一套分割参数得到结果的信息提取精度, 因此主要通过目标函数来评价分割后的影像对象的信息量。本课题通过分析滑坡对象内同质性和对象间异质性指标来确定目标函数, 提出一种基于 SSC (Scale-Shape-Compactness) 的滑坡对象构建方法。该方法针对面向对象信息提取多尺度分割中分割参数设置问题, 通过结合影像对象内均质度、对象间异质度以及对象面积创建的分割质量评价指标, 分别从多尺度分割的尺度 (Scale) 参数、形状因子 (Shape) 参数和紧致度 (Compactness) 参数三方面进行优化, 以得到面向对象遥感滑坡信息的最优分割参数。

(2) 面向对象滑坡信息提取特征选择方法的研究

面向对象影像分析技术使得高分辨率遥感影像的特征维数从多波段的几维扩展到由光谱、形状、纹理和语义信息等特征组成的数百维。随机森林以其适合处理小样本大数据量的优势, 在面向对象特征选择应用中表现出了精度高、速度快、稳定性好的特点。本课题分析随机森林算法的原理, 提出一种基于随机森林的面向对象滑坡特征选择方法。该方法针对面向对象遥感滑坡信息提取中特征选择问题, 结合随机森林在特征选择方面的优势, 通过 R 语言编写的程序, 分别对遥感滑坡信息提取中各类地物进行特征选择运算, 以得到目标地物的诊断性特征。

(3) 滑坡信息提取中 DEM 空间分辨率的选择

数字高程模型 (DEM) 是识别滑坡的重要特征, 而在实际应用中通常难以获取研究区的高分辨率 DEM 数据。本课题将原始分辨率 DEM 重采样为不同空间分辨率, 采用 DEM 及其衍生数据辅助高分遥感影像的面向对象滑坡信息提取方法进行实验, 分析 DEM 分辨率对滑坡提取的影响并确定满足应用的分辨率要求。

1.3.2 技术路线

(1) 影像分割和特征运算

遥感影像分割主要由对象内同质性和对象间异质性来决定, 由于无法获取每一套分割参数得到结果的信息提取精度, 因此主要通过目标函数来评价分割后的影像对象的信息量。本课题通过结合对象内标准差和莫兰 I 指数确定目标函数, 利用间隔尺度分割得到的目标函数值拟合平稳目标函数, 由目标函数的极值点确定最优的分割尺度参数。

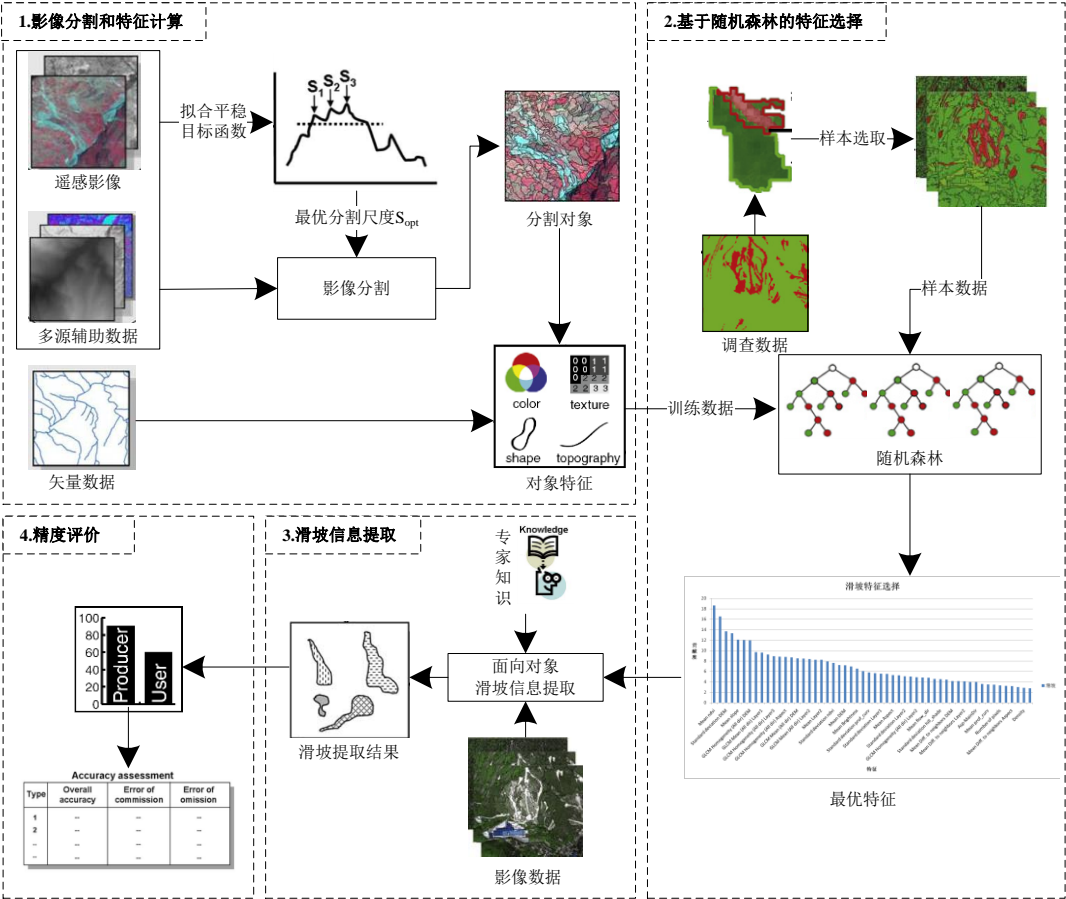


图 1-1 技术路线图

(2) 基于随机森林的特征选择

利用最优尺度参数进行遥感影像分割，将得到的影像对象计算光谱、形状、纹理等影像特征作为随机森林特征选择的训练数据，将部分地面调查得到的滑坡矢量数据作为作为随机森林的样本数据。根据对特征贡献度的计算获取对分类精度贡献最大的特征组合。

(3) 滑坡信息提取

利用随机森林方法选择出的最优特征结合影像数据和专家知识，通过面向对象滑坡信息提取方法，对遥感数据进行滑坡信息提取，得到滑坡提取结果。

(4) 精度评价

为了得到面向对象遥感滑坡信息提取的精度，我们以原始的滑坡地面调查数据作为验证数据对滑坡信息提取结果进行精度评价。

(5) DEM 空间分辨率影像分析

为分析 DEM 分辨率对滑坡提取的影响并确定满足应用的分辨率要求，将原始 10m 分辨率 DEM 重采样为空间分辨率不同的 15m、20m、30m、50m、75m、100m，采用 DEM 及其衍生数据辅助高分遥感影像的面向对象滑坡信息提取方法进行实验。

§1.4 论文章节内容安排

论文章节内容主要安排如下：

第一章，绪论。简明扼要地阐述了本文的研究背景及意义，主要对遥感滑坡信息提取现状、最优分割参数选择现状和信息提取特征选择现状进行了介绍，对文章的研究内容及技术路线进行了规划，并对论文各个章节的内容进行了部署。

第二章，面向对象遥感滑坡信息提取方法。从三个方面介绍面向对象影像分析的关键步骤，并详细阐述了面向对象影像分析应用于遥感滑坡信息提取的技术流程。

第三章，基于 SSC 的滑坡对象构建方法。首先介绍多尺度分割的概念、算法和技术流程，然后提出 SSC（Scale-Shape-Compactness）多尺度分割参数优化方法，最后进行基于 SSC 的滑坡对象构建实验。

第四章，随机森林滑坡特征选择方法。首先从决策树和随机森林两个方面介绍随机森林的基本原理，然后提出基于对象的随机森林滑坡特征选择方法，最后进行随机森林滑坡特征选择实验。

第五章，面向对象滑坡信息提取中 DEM 空间分辨率影响分析。本章是对第二、三、四章的综合应用，同时，其结论用于滑坡信息提取中指导 DEM 产品的选择。

第六章，总结与展望。总结了全文的研究结果，并在此基础上对面向对象滑坡信息提取技术的研究提出了展望。

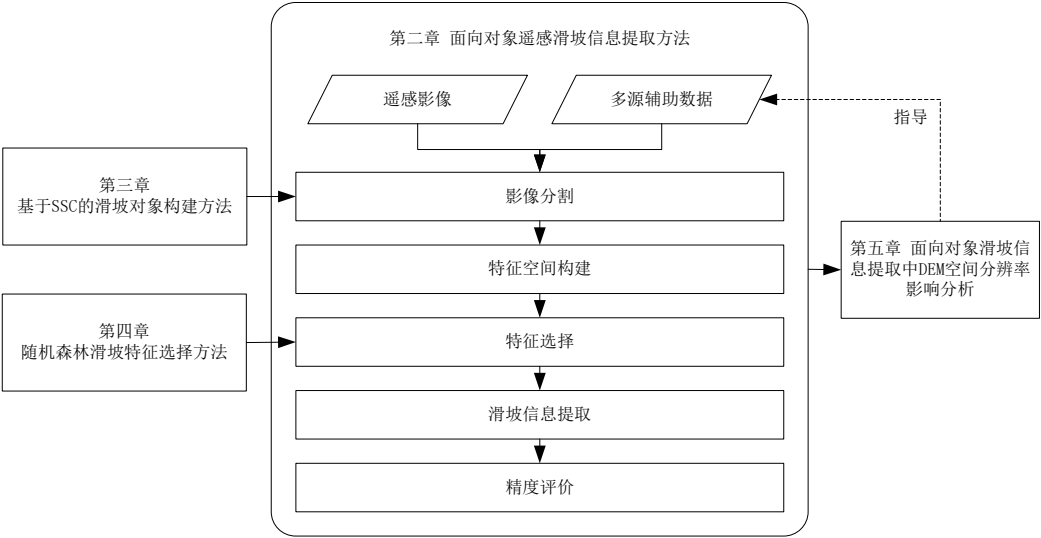


图 1-2 章节关系图

第二章 面向对象滑坡信息提取方法

目前比较常用的面向对象影像分析工具有美国 Trimble Navigation 公司的 eCognition（原为德国 Definiens 公司产品）、美国 Textron Systems 公司的 Feature Analyst、美国 ERDAS 公司的 ERDAS IMAGINE Objective 面向对象的信息提取模块、美国 Exelis Visual Information Solutions 公司的 ENVI EX（Feature Extraction）面向对象空间特征提取模块以及加拿大 PCI Geomatics 公司 PCI Geomatica 软件产品的 Feature ObjEX 智能地物要素提取模块。本文主要采用 eCognition Developer 8.7 作为面向对象滑坡信息提取的处理工具。

§2.1 面向对象影像分析技术

2.1.1 影像分割

在描述感兴趣地物诊断性特征之前，一个重要的步骤就是创建能包含这类地物的影像对象。这通过影像分割来实现，它是根据像元值的同质性将影像分成不同的对象或区域的一种处理方法。影像分割能通过不同的方式实现，例如密度分割、分裂和合并等技术。分割算法用于在像素级别细分整个影像，或者将其他域的特定影像对象分成更小的影像对象。eCognition 软件提供几种不同的方法来进行影像分割，只要你想基于影像波段信息创建新的对象层，分割算法就是必需的。这些分割算法也是完善现有影像对象的一种非常有价值的工具，它们通过将影像对象分成更小的部分来进行更详细的分析。eCognition 软件提供影像分割方法主要有：棋盘分割、四叉树分割、对比分裂分割、多尺度分割、光谱差异分割、多阈值分割和对比度过滤器分割等^[51]。以下是对几种分割方法的详细介绍：

（1）棋盘分割

通过设置相关参数，棋盘分割算法将像素域或影像对象域分割成大小相同的正方形影像对象。将边界对齐于影像边框且大小固定的正方形网格应用到区域内的所有对象，使得每一个对象都是沿着这些个网线切割。

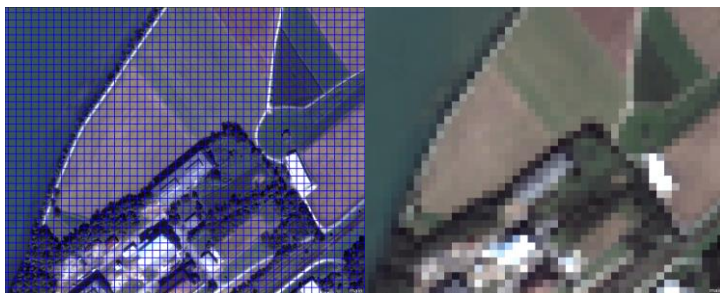


图 2-1 棋盘分割结果（对象大小：20）

如果在棋盘分割中将对象大小值设为 20，那么分割后的每个影像对象的大小都是 20×20 个像素，分割结果如图 2-1。

(2) 四叉树分割

四叉树分割算法将像素域或影像对象域分割成由正方形对象组成的四叉树格网。那些组成四叉树格网的正方形，每个边长都是 2 的幂并且边界对齐于影像边框。它们适用于影像区域内的所有对象，每个对象都是沿着这些网格线切割。四叉树结构是建立让每个正方形具有最大可能的大小，同时也符合由模型和尺度参数确定的同质性准则。其中，最大的正方形对象大小为 256×256 ，也即是 65536 个像素。该算法通常适用于影像背景和前景差别比较明显的情况下，可以把大量的不感兴趣的区域先分离出来。

如果在四叉树分割中模式设为颜色，尺度设为 100，则分割结果如图 2-2。

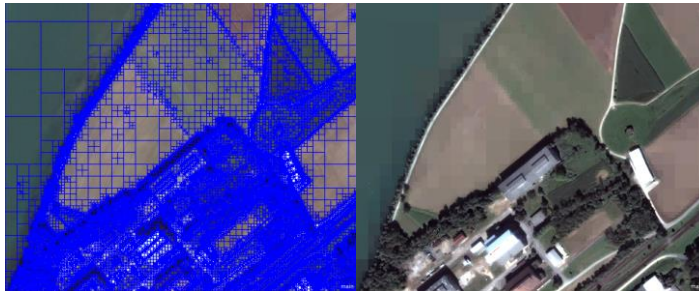


图 2-2 四叉树分割结果（模式：颜色，尺度：100）

(3) 对比分裂分割

对比分裂分割算法能将影像或影像对象分割成亮和暗两种区域。它主要基于一个阈值，该阈值能最大化亮对象（由像素值大于阈值的像素组成）与暗对象（由像素值小于阈值的像素组成）之间的对比度。该算法在影像对象域中分别为每个影像对象估计最优阈值。如果在影像对象域中选择像素层，那么该算法首先执行一次棋盘分割，然后再分别对每个正方形对象进行进一步的分割。算法通过考虑不同像素值作为潜在阈值来实现最优化。作为测试的阈值范围从最小阈值到最大阈值，而中间值则是根据步长大小和步长类型参数来选择。如果测试的阈值满足最小暗区域与最小亮区域的准则，那么就需要评价亮对象和暗对象的对比度。得到最大对比度的测试阈值将作为最优阈值，同时也将其应用于分割当中。

(4) 多尺度分割

多尺度分割算法是目前最常用的分割算法，它对于给定分辨率的影像对象，最大限度地减少影像对象的局部异质度。算法可以在现有影像对象层或者像素层的基础上创建新的影像对象层。多尺度分割算法连续地合并像素或现有的影像对象，因此，它是一个基于相邻区域合并的自下而上的分割算法。多尺度分割是一个优化过程，对于给定数量的影像对象，它最大限度的减小平均异质性，并且最大化各自的同质性。

合适的分割参数能够有效地确保目标地物边界提取的准确性，只有正确地描绘目标地物的边界，才能准确地进行影像信息提取。该方法将在本文的第 3.1 节做详细介绍。

如果在多尺度分割中尺度设为 30，形状指数设为 0.1，紧致度设为 0.5，则分割结果如图 2-3。



图 2-3 多尺度分割结果（尺度：30，形状指数：0.1，紧致度：0.5）

（5）光谱差异分割

光谱差异分割算法根据影像层的强度均值合并相邻的影像对象。如果相邻影像对象强度均值之间的差异小于所给定的最大光谱差异值，那么这两个相邻的影像对象将被合并。该分割算法的主要目的是对光谱差异小于阈值的相邻对象进行合并，通过合并之前分割产生的光谱相似的影像对象，来改进现有的分割结果。它不能通过像素域创建新的影像对象层。

（6）多阈值分割

多阈值分割算法是基于像素值来分割影像对象域。该算法基于用户设定的阈值，创建影像对象并将它们分类。它也可以基于像素值创建未分类的影像对象。

（7）对比度过滤器分割

像素过滤器通过对比度和梯度来检测潜在对象，并且创建合适的原始对象。通过集成的重塑操作修改影像对象的形状，以帮助形成连贯和紧凑的影像对象。像素分类的结果被存储在内部专题层。每个像素被分类为以下类别之一：1）没有对象；2）对象在第一层；3）对象在第二层；4）对象在这两个层；5）通过阈值忽略。最后通过棋盘分割将这些专题图层转换为影像对象层。作为分割算法的第一步，对比度过滤器分割对整个影像的分析性能有了很大的改进。

2.1.2 特征空间构建

影像分割构建的影像对象是面向对象遥感影像分析中最为明显的优势，它们不仅仅只有光谱特征，还包含了形状、纹理上下文关系等特征^[52]。eCognition 软件提供特征类型主要有两类：1）对象特征，指影像对象的属性，例如面积。2）全局特征，该特征并不关联到一个独立的影像对象，例如一个指定类别的影像对象数量。对象特征是通过评估影像对象本身和它们嵌入的影像对象层次而计算得到的。对象特征主要分为：自定义特征、图层值特征、几何特征、位置特征、纹理特征、对象变量特征、层次特征和专题属性特征等^[51]，如图 2-4。下面将分别介绍这几种类特征：

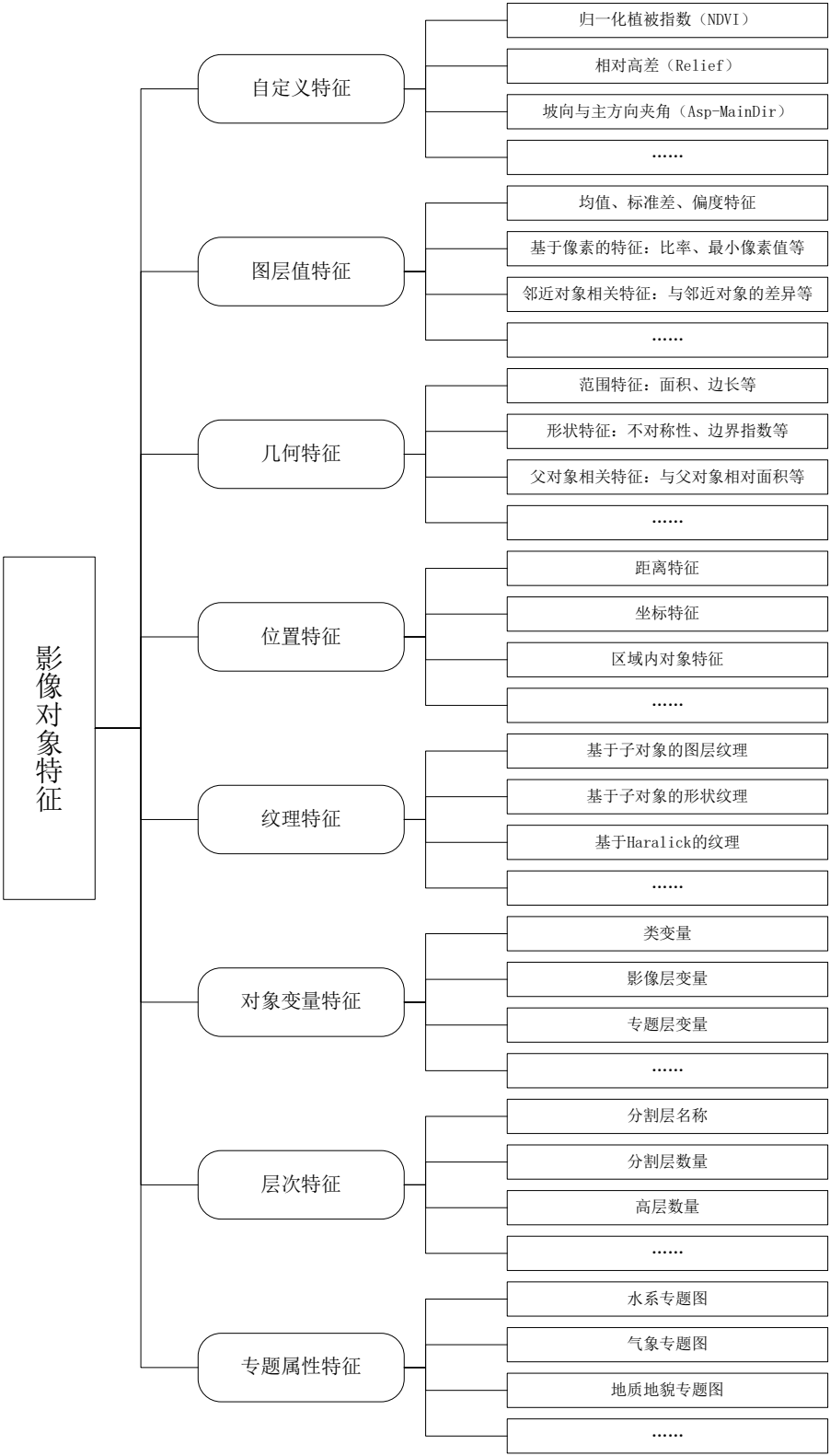


图 2-4 影像对象特征

(1) 自定义特征

自定义特征是用户根据实际应用需求, 利用已有特征的算数或关系运算而自主创建的具有特殊意义的特征。自定义特征在面向对象滑坡信息提取中有着非常重要的作用, 可用于计算归一化植被指数 (NDVI)、相对高差 (Relief) 和坡向与影像对象主轴方向的夹角 (Asp-MainDir) 等特征。

(2) 图层值特征

图层值特征用来评价影像对象像素值的第一、第二和第三统计矩 (均值、标准差和偏度) 以及与这些对象相关联的其他影像对象像素值。这些特征用来描述那些光谱衍生信息的影像对象。图层值特征包括: 比如均值、标准差、偏度、比率、基于像素的特征 (比率、最小像素值、最大像素值等)、邻近对象相关特征 (与邻近对象的差异等)、父对象相关特征 (与父对象的差异、与父对象的比率等)、场景相关特征 (与场景的差异、与场景的比率等) 以及色调、饱和度、亮度特征等。

(3) 几何特征

几何特征用来评价影像对象的形状。基本几何特征是基于影像对象中的像素来计算得到的。因为影像是基于栅格的, 所以几何特征可能是旋转变量 (在影像对象旋转之后会出现不同的特征值)。如果某一类影像对象因为它们的形状而比较显著, 那么就可以找到描述它们的几何特征。几何特征包括: 范围特征 (面积、边长、长宽比等)、形状特征 (不对称性、边界指数、紧致度等)、父对象相关特征 (父对象的相对面积等)、基于多边形的特征 (多边形平均边长、多边形周长等)、基于骨架的特征 (平均分支长度、Delaunay 三角形的数量等) 等。

(4) 位置特征

位置特征是指影像对象相对于整个场景的位置, 是像素坐标的定义。当运用于地理空间数据时, 这些特征会受到特别的关注, 因为影像对象可以由它的地理位置来描述。

(5) 纹理特征

纹理特征被用来评价影像对象中的纹理。它们包括那些基于子对象分析的纹理特征, 这将有助于评价高纹理数据。此外, 基于 Haralick 灰度共生矩阵的纹理特征也是可用的。

(6) 对象变量特征

对象变量是局部变量, 相对于场景变量, 它们为每个独立对象存储变量值 (可以看成是每个影像对象有属于它自己版本的变量)。项目中的每个影像对象都有一个实例, 如类变量、影像层变量、专题层变量、分割层变量、特征变量等。

(7) 层次特征

层次特征提供关于影像对象层次结构中影像对象的嵌入信息。

(8) 专题属性特征

专题属性特征被用来描述那些具有专题图层信息的影像对象。如果场景中包含专题图层, 那么它的专题对象属性可被用来构建分类规则。依靠专题层的属性, 我们可以应用更广泛的不同特征, 如水系专题图、气象专题图、地质地貌专题图等。

2.1.3 特征选择

根据模式识别相关理论^[53]，结合遥感影像分类的相关技术，本节针对特征选择的基本概念，以及面向对象分类中的特征选择进行简单的阐述，具体特征选择的研究内容在第 4 章进行详细讨论。

2.1.3.1 特征选择的定义

- 按照对分类的作用，特征可以分为以下两类：
- 1. 相关特征（Relevant Feature）。指对分类作用相对较大的特征，若这类特征不参与分类，则会降低分类精度；
 - 2. 无关特征（Irrelevant Feature）^[53]。指在分类中起到反作用的特征，即使分类中不采用这类特征，分类精度也不会受到太大影响，甚至可以提高分类精度，分类的速度将会得到提高。

不同的研究者从不同的角度对特征选择进行了定义^{[54][55]}。特征选择可以按照作用的不同定义为：从特征维度为 D 的一组原始特征空间中，根据某种特征评估准则，选择出 $d(d < D)$ 个重要特征，使得特征空间的维度减少，剔除类间差别小的特征，保留类间差别大的特征，以实现特征空间结构简化的目的。

从特征选择的定义中我们可以看出，特征选择是在保留原特征集全部或大部分类别信息的条件下，利用样本集内部的信息，从候选特征集当中选择一个最优特征子集的过程。换句话说，特征选择的作用就是为了提高信息提取分类器的分类性能，降低特征空间维度，而从候选特征中选出最有效的特征。

2.1.3.2 特征选择的分类

特征的搜索策略和评价准则是特征选择过程中的两个重要内容。接下来将从这两个方面对特征选择进行分类：

（1）按特征搜索策略分类

按特征搜索策略，我们可以将特征选择算法可分为：完全式搜索特征选择算法、启发式搜索特征选择算法以及随机式搜索特征选择算法。详细描述见图 2-5 和表 2-1 所示：

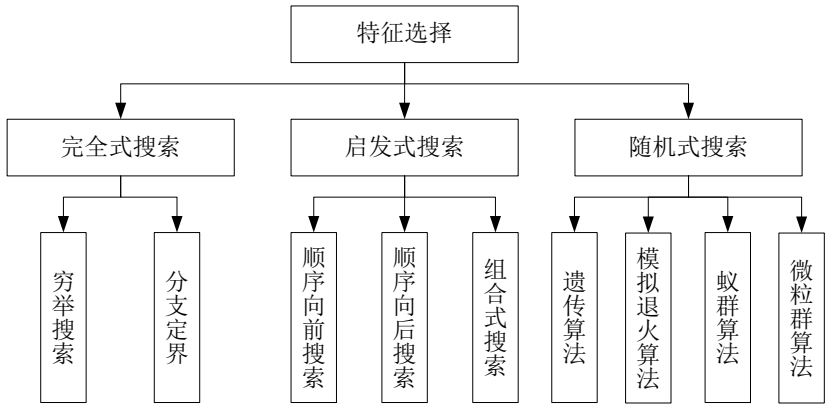


图 2-5 按搜索策略分类的特征选择算法图

表 2-1 按搜索策略分类的特征选择算法表

特征选择算法类别	描述
完全式搜索 (Complete Search)	完全式搜索算法通常能够得到最优的特征集，但因为效率非常低，因此在实际应用中很少被采纳。常用的完全式搜索算法包括：穷举搜索算法和分支定界算法等。
启发式搜索 (Heuristic Search)	启发式搜索算法在计算复杂性和搜索最优性之间做了很好平衡，通常该方法可以去掉部分不重要的无关特征。常用的启发式搜索算法包括：顺序搜索、最优特征组合法。
随机式搜索 (Random Search)	随机式搜索算法可以随机的生成一个特征子集，该特征子集与当前特征集是无关的。常用的随机搜索算法包括：LVF (Las Vegas Filter) ^[56] 、遗传算法、模拟退火算法、蚁群算法和微粒群算法等。

（2）按评价准则分类

按评价准则的不同，我们可以将特征选择算法可分为：过滤式特征选择算法、封装式特征选择算法和集成式特征选择算法。详细描述见图 2-6 和表 2-2 所示：

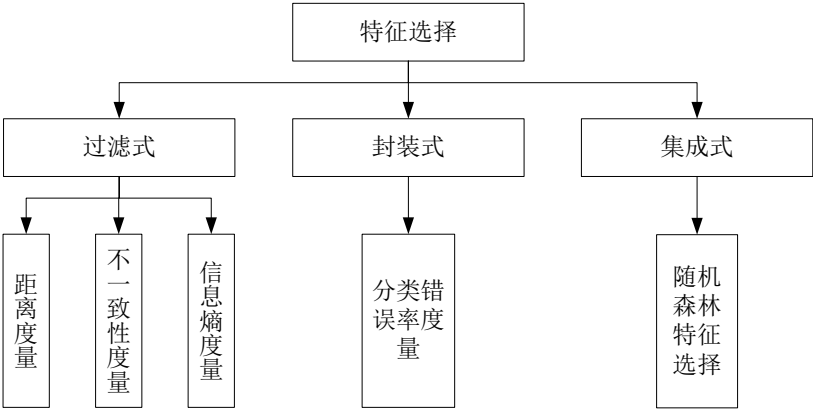


图 2-6 按评价准则的特征选择分类

表 2-2 按评价准则分类的特征选择算法表

特征选择算法类别	描述
过滤式 (Filter)	过滤式特征选择算法通常不以分类精度作为评价韩式，它直接对训练样本数据的内在特征进行评估，因此，该算法的效率较高。常用的过滤式特征选择算法包括：距离度量、不一致性度量、信息熵度量等。
封装式 (Wrapper)	封装式特征选择算法，通过事先确定学习算法，根据分类器的错误概率选择特征集，该特征集使得分类器的错误概率最小。相对于过滤式特征选择算法，该算法需要在评价的过程中多次调用学习算法，因此通常情况下消耗的时间会比较多。
集成式 (Integrator)	相比较于过滤式和封装式特征选择算法而言，集成式特征选择算法的定义则并不是那么的明确。例如，随机森林特征选择算法就是一种集成式的方法，它是由多棵决策树组合而成的，具有同样的基分类器。虽然集成式的特征选择算法组合的方式多种多样，但是，在任意组合后的效果并不一定会由于它的基分类器。尽管如此，集成式的特征选择方法任然是目前特征选择的热点研究方向。

2.1.3.3 面向对象信息提取的特征选择

通过 2.1.2 节不难看出, 相对于单个像素而言, 影像对象的特征信息更为丰富, 更接近实际地物。而伴随着特征信息的多样化, 特征维数也随之增加, 若不对所有特征加以有效选择, 必将导致“Hughes”现象^[57]。由此可见特征选择是面向对象信息提取中不可或缺的一步。研究表明, 面向对象信息提取分类算法与参与的特征之间的并不是简单的线性关系^[58]。所以, 特征选择是面向对象滑坡遥感信息提取中需要重点解决的关键问题之一^[59]。

本文针对以上问题, 提出了基于随机森林的面向对象滑坡特征选择方法。该方法针对面向对象遥感滑坡信息提取中特征选择问题, 结合随机森林在特征选择方面的优势, 通过 R 语言编写的程序, 分别对遥感滑坡信息提取中各类地物进行特征选择运算, 以得到目标地物的诊断性特征, 具体研究内容在 4 章详细阐述。

§2.2 遥感影像滑坡信息提取流程

面向对象影像分析是一种知识驱动的信息提取方法, 它基于专家知识结合地物的光谱、形态学和上下文等诊断性特征精确地提取地物信息。然而, 专题遥感信息提取通常需要与特定应用相联系, 因此, 本文将面向对象影像分析技术与滑坡地物特征相结合, 提出了一种面向对象滑坡信息提取技术流程。整个滑坡信息提取流程(图 2-7)主要分为三个阶段: 第一个是数据准备阶段, 主要是准备实验中将用到的遥感数据, 包括多光谱数据、NDVI 数据、DEM、坡度、坡向、地表曲率、山体阴影、水系图等; 第二个是数据处理阶段, 主要对前面提供的遥感数据进行处理分析, 利用面向对象分析的方法提取滑坡信息, 包括影像分割、滑坡识别、滑坡分类和信息提取结果导出等过程; 第三个是精度评价阶段, 主要利用野外实地调查结果作为评价标准, 对自动提取的结果进行精度评价, 并将最终的结果编录与制图。

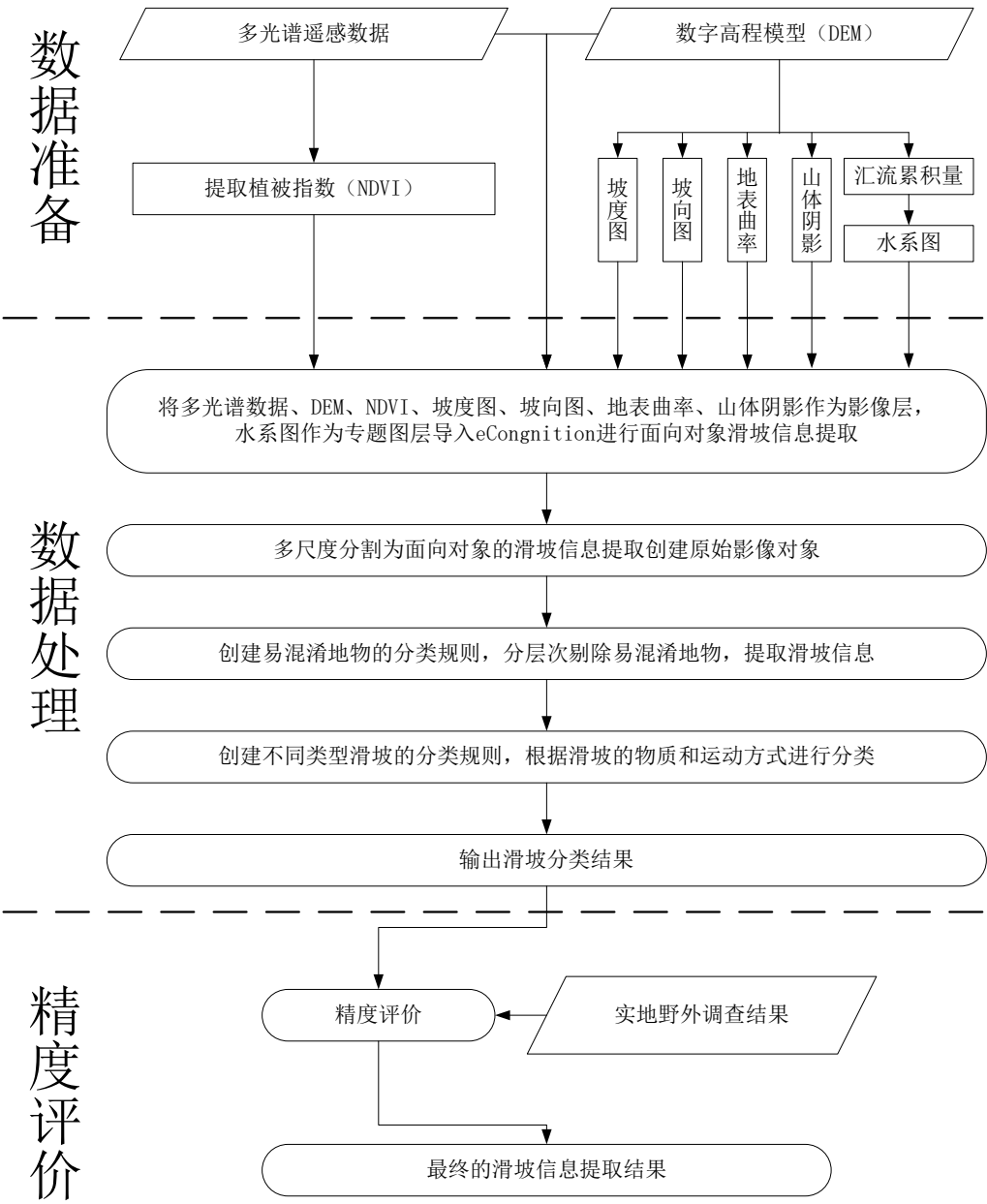


图 2-7 遥感影像滑坡信息提取流程

§2.3 本章小结

本章主要介绍面向对象滑坡信息提取中涉及到的基本方法和技术。从三个方面介绍面向对象影像分析的关键步骤，并详细阐述了面向对象影像分析应用于遥感滑坡信息提取的技术流程。本章介绍的方法和技术是本文后续研究重要的理论基础。

第三章 基于 SSC 的滑坡对象构建方法

影像对象是面向对象遥感信息提取的基本单元，如何构建影像对象是进行面向对象遥感滑坡信息提取的基础，而影像分割则是产生影像对象的有效方法，因此影像分割的好坏直接决定了面向对象信息提取的准确性。多尺度分割是目前公认最好的遥感影像分割算法，其分割效果主要取决于分割参数的设置，分割参数包括：分割尺度、颜色因子、形状因子、光滑度和紧致度等。如何设置最优的分割参数，使得多尺度分割生成的多边形影像对象能够最有效的表达影像中的目标地物，是文章接下来要讨论的重点内容。本章首先介绍多尺度分割的基本原理，然后介绍 SSC (Scale-Shape-Compactness) 多尺度分割参数优化方法，之后在此基础上进行基于 SSC 的滑坡对象构建实验，最后对基于 SSC 的滑坡对象构建方法进行总结。

§3.1 多尺度分割的基本原理

多尺度分割算法能够有效地确保目标地物边界提取的准确性，具有不依赖先验知识来进行影像分割的潜力^[45]。在面向对象滑坡信息提取研究中，将主要采用多尺度分割算法进行遥感影像的分割操作。文章的 2.2.1 小节已经对该方法有了简要的说明，下面将详细介绍面向对象多尺度分割技术。

多尺度分割算法通常采用异质性最小的 FNEA 方法，算法的核心思想是，将单个像素合并为较小的影像对象，再合并影像对象形成较大的多边形对象^[60]。为了使整幅影像的所有对象的平均异质性最小，在进行分割之前需要确定光谱因子和形状因子，这两种因子是影响异质性的关键因子^[61]。

几种异质性度量具体介绍如下：

(1) 光谱异质性

光谱异质性参量 (h_{color}) 表示如下：

$$h_{color} = \sum_c \omega_c \cdot \sigma_c \quad (3-1)$$

在公式 3-1 中，波段数用 c 表示，波段的灰度值用 σ_c 表示，影像层的权重用 ω_c 表示。

如果用 n_{obj1} , σ_c^{obj1} , n_{obj2} , σ_c^{obj2} ，分别表示的是两个相邻区域的面积和方差，那么我们可以将合并后的光谱异质性度量准则用公式 3-2 来表示：

$$h_{color} = \sum_c \omega_c (n_{Merge} \cdot \sigma_c^{Merge} - (n_{obj1} \cdot \sigma_c^{obj1} + n_{obj2} \cdot \sigma_c^{obj2})) \quad (3-2)$$

在上式中合并后区域的面积和方差分别用 n_{Merge} 、 σ_c^{Merge} 表示，要参与分割合并波段的权重用 ω_c 来表示。

(2) 形状异质性

形状异质性可以由光滑度参数 h_{sm} 和紧致度参数 h_{cm} 来反映形状特征。合并区域的光滑程度由光滑度参数来反映，而合并后的紧凑度则由紧致度参数来体现。这两个参数可以表达为：

$$h_{sm} = \frac{l}{b} \quad h_{cm} = \frac{l}{\sqrt{n}} \quad (3-3)$$

公式 3-3 中 l 、 n 、 b 分别表示的是区域的周长、面积和最小外接矩形的周长。如果我们用 l_{obj1} 、 b_{obj1} 、 l_{obj2} 、 b_{obj2} 来表示两个相邻区域的形状，用 h_{smooth} 和 $h_{compact}$ 来表示合并后的形状参数，那么可以得到这两个形状差异性度量的准则：

$$h_{smooth} = n_{Merge} \cdot \frac{l_{Merge}}{b_{Merge}} \cdot (n_{obj1} \cdot \frac{l_{obj1}}{b_{obj1}} + n_{obj2} \cdot \frac{l_{obj2}}{b_{obj2}}) \quad (3-4)$$

$$h_{compact} = n_{Merge} \cdot \frac{l_{Merge}}{\sqrt{n_{Merge}}} \cdot (n_{obj1} \cdot \frac{l_{obj1}}{\sqrt{n_{obj1}}} + n_{obj2} \cdot \frac{l_{obj2}}{\sqrt{n_{obj2}}}) \quad (3-5)$$

在公式 3-4 和 3-5 中，我们将区域合并后的面积、最小外接矩形的周长以及周长分别用 n_{Merge} 、 b_{Merge} 、 l_{Merge} 来表示。

而形状异质性可以用前面提到的光滑度参数和紧致度参数来表示：

$$h_{shape} = \omega_{compact} \cdot h_{compact} + (1 - \omega_{compact}) \cdot h_{smooth} \quad (3-6)$$

在公式 3-6 中，在计算形状异质性度量准则中，紧致度参数所占的权重用 $\omega_{compact}$ 来表示。

(3) 综合异质性度量准则

对上述的光谱异质性参数和形状异质性参数度量准则进行综合，我们可以采用公式 3-7 来表示综合异质性度量准则：

$$f = \omega \cdot h_{color} + (1 - \omega) \cdot h_{shape} \quad (3-7)$$

在上式中，光谱异质性度量准则和形状的形状异质性度量准则分别用 h_{color} 和 h_{shape} 来表示，

而光谱异质性度量准则在综合异质性度量准则中所占的比重则用 ω 表示。

在分割时我们必须遵循两个基本原则：首先尽可能大的设置光谱因子的权重；其次设置合适的形状因子，使得分割生成的影像对象平滑度和紧致度达到最佳^[60]。

§3.2 SSC 多尺度分割参数优化方法

3.2.1 分割质量评价

分割质量评价主要有 3 种思路：（1）对分割算法本身进行评价的分析法；（2）对分割结果进行评价，选择最优结果的优度实验法；（3）将分割结果与已知参考结果进行比较的差异实验法。综合以上 3 种质量评价方法，虽然（2）优度实验法要对所有可能的分割结果进行穷举选择，但该方法具有更好的可行性，分割结果通过量化评价指标，不仅与分割算法内部结构特征无关，而且具有客观性和普适性。因此本文采用优度实验法对分割结果进行评价。

理想的影像分割结果，其评价准则通常包括：（1）对象内部各像素之间具有较好的同质性，保证分割产生的影像对象尽量是纯净对象；（2）分割得到的对象应该与邻接对象高度可分，具有良好的异质性；（3）对象面积大小对分割质量评价结果有不同的权重影响。

为了客观地找到最优分割尺度，我们参考了 Espindola 等在 2006 年提出目标函数来评价遥感影像的分割质量^[62]，该目标函数主要由对象内同质性和对象间异质性这两个部分组成。

（1）对象内同质性

对象的同质性通过利用对象内方差（ V ）来表示，其计算公式如下：

$$V = \frac{\sum_{i=1}^n a_i v_i}{\sum_{i=1}^n a_i} \quad (3-8)$$

其中， n 表示研究区分割对象的总数， a_i 表示对象 i 的面积， v_i 表示对象 i 的方差。

影像同质性为对象内方差的加权平均结果，权重是各个对象的面积。为了避免小面积对象对分割结果造成影响，该方法将大面积对象赋予更多的权重。对象内方差 V 越小，则影像对象内各像素之间的异质性越小，也即表示对象内具有较高的同质性，可以得到理想的影像分割结果。

（2）对象间异质性

Moran's I 指数是评价空间自相关性的重要指标，为了使分割得到的对象与邻接对象高度可分，影像对象之间的异质性评价指标采用 Moran's I 指数（ I ）来表示。其计算公式如下：

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (x_i - \bar{X})(x_j - \bar{X})}{\{\sum_{i=1}^n (x_i - \bar{X})^2\} (\sum_{i=1}^n \sum_{j=1}^n w_{ij})} \quad (3-9)$$

其中, n 表示研究区分割影像对象的总数; w_{ij} 表示影像对象 i 与影像对象 j 的空间邻近性, 当影像对象 i 与影像对象 j 相互邻接时, 则 w_{ij} 值为 1, 若不互相邻接则 w_{ij} 值为 0;

x_i 与 x_j 分别表示影像对象 i 与影像对象 j 的平均灰度值; \bar{X} 表示整个影像的平均灰度值。

Moran's I 指数反应了对象的均值与相邻对象均值的差异, I 值越小, 影像对象空间自相关性越低, 即影像对象与邻接对象高度可分, 可以得到理想的影像分割结果。

(3) 分割质量评价函数

为了让对象内同质性和对象间异质性有相同的贡献度, 需要对它们进行归一化:

$$F(V) = \frac{V - V_{\min}}{V_{\max} - V_{\min}} \quad F(I) = \frac{I - I_{\min}}{I_{\max} - I_{\min}} \quad (3-10)$$

分割质量评价函数为:

$$F(V, I) = F(V) + F(I) \quad (3-11)$$

$F(V)$ 表示对象内同质性, $F(I)$ 表示对象间异质性, $F(V, I)$ 为遥感影像的分割质量综合评价函数。

(4) 平稳目标函数

目标函数的极大值是最优影像分割尺度的一个统计指标^[62]。然而, 一个单一的最优尺度是不足以解决影像的空间结构和景观结构之间的关系。为了获得复合最优尺度作为替代, 需要创建了一个简单的平稳目标函数

$$F(\text{plateau}) = F(V, I)_{\max} - \sigma \quad (3-12)$$

其中, $F(V, I)_{\max}$ 表示极大值, σ 表示目标函数的标准差。

平稳目标函数值是用于划分绘图比例因子和目标函数在 x 轴和 y 轴分别创建的曲线在平稳时期的下边界。这个平稳目标函数的假设是峰值接近目标函数的最大值, 因此, 欠分割和过分割之间的平衡仍然存在。此外, 该峰值相对于它们的近邻区域而言彼此明显不同且为局部最优。由于感兴趣的目标对象占据遥感影像的主要部分, 因此这类平稳模式通常是可预见的。

3.2.2 SSC 分割参数优化方法

分割结果的好坏主要取决于分割参数的设置,多尺度分割的参数主要包括:分割尺度、颜色因子、形状因子、光滑度和紧致度。其中,颜色因子和形状因子的权重之和为 1,如果颜色因子设为 0.6,那么形状因子的为 0.4。而形状因子又是由光滑度和紧致度组成,光滑度和紧致度权重之和为 1,设置了紧致度,相对地光滑度权重也就被确定了。

在 eCognition 软件的多尺度分割过程中,仅需要设置尺度 (Scale)、形状因子 (Shape) 和紧致度 (Compactness) 这 3 个参数,颜色因子 (Color) 和光滑度 (Smoothness) 可以通过形状因子和紧致度计算得到。针对多尺度分割中的尺度、形状因子和紧致度参数,结合分割质量评价函数,提出 SSC (Scale-Shape-Compactness) 多尺度分割参数优化方法。方法包括以下步骤:

- 1) 预设不变参数值。首先计算尺度参数的优化值,此时,形状因子和紧致度分别预设为默认值 0.1 和 0.5;
- 2) 参数值间隔采样。定义尺度参数的范围和采样间隔,生成尺度样本参数值;
- 3) 多尺度分割。将尺度样本参数值和预设的形状因子与紧致度参数带入多尺度分割,得到各自的分割结果;
- 4) 计算分割质量评价值。计算每个分割结果对应的分割质量评价值;
- 5) 拟合目标函数曲线。以尺度参数作为横坐标,分割质量评价值作为纵坐标,拟合目标函数曲线;
- 6) 计算目标函数的极大值。根据拟合的目标函数曲线,计算出目标函数的极大值;
- 7) 计算平稳目标函数。根据目标函数极大值和样本的标准差,得到平稳目标函数;
- 8) 得到优化参数。通常情况下,选择最大极大值作为优化参数值。特殊情况时,考虑选择大于平稳目标函数的其它极大值作为优化参数值。
- 9) 参照 1-8 步,计算其它参数的优化值。若已经计算出优化值,则将优化值作为预设值,参照 1-8 步,计算出剩下参数的优化值。

§3.3 基于 SSC 的滑坡对象构建实验

3.3.1 实验区概况及数据介绍

研究区选在湖南省资兴市东部的东江水库地区 (如图 3-1),地理范围为 $113^{\circ}23'36'' \sim 113^{\circ}35'24''\text{E}$, $25^{\circ}54'47'' \sim 26^{\circ}3'21''\text{N}$,遥感数据为 2006 年 12 月 17 日获取的 QuickBird2 多光谱数据。研究区地处罗霄山脉的西南段,属湘江上游支流耒水流域,为山地丘陵地貌,主要山脉有八面山、七宝山及瑶岗仙山等,出露地层主要为砂岩、灰岩、砂质板岩及易于风化的中粗粒花岗岩。2006 年 7 月 15 日,由于受到强热带风暴碧利斯的影响,该地区发生了严重的洪涝灾害,并诱发山体出现大面积的、连续性的塌方、滑坡和泥石流。

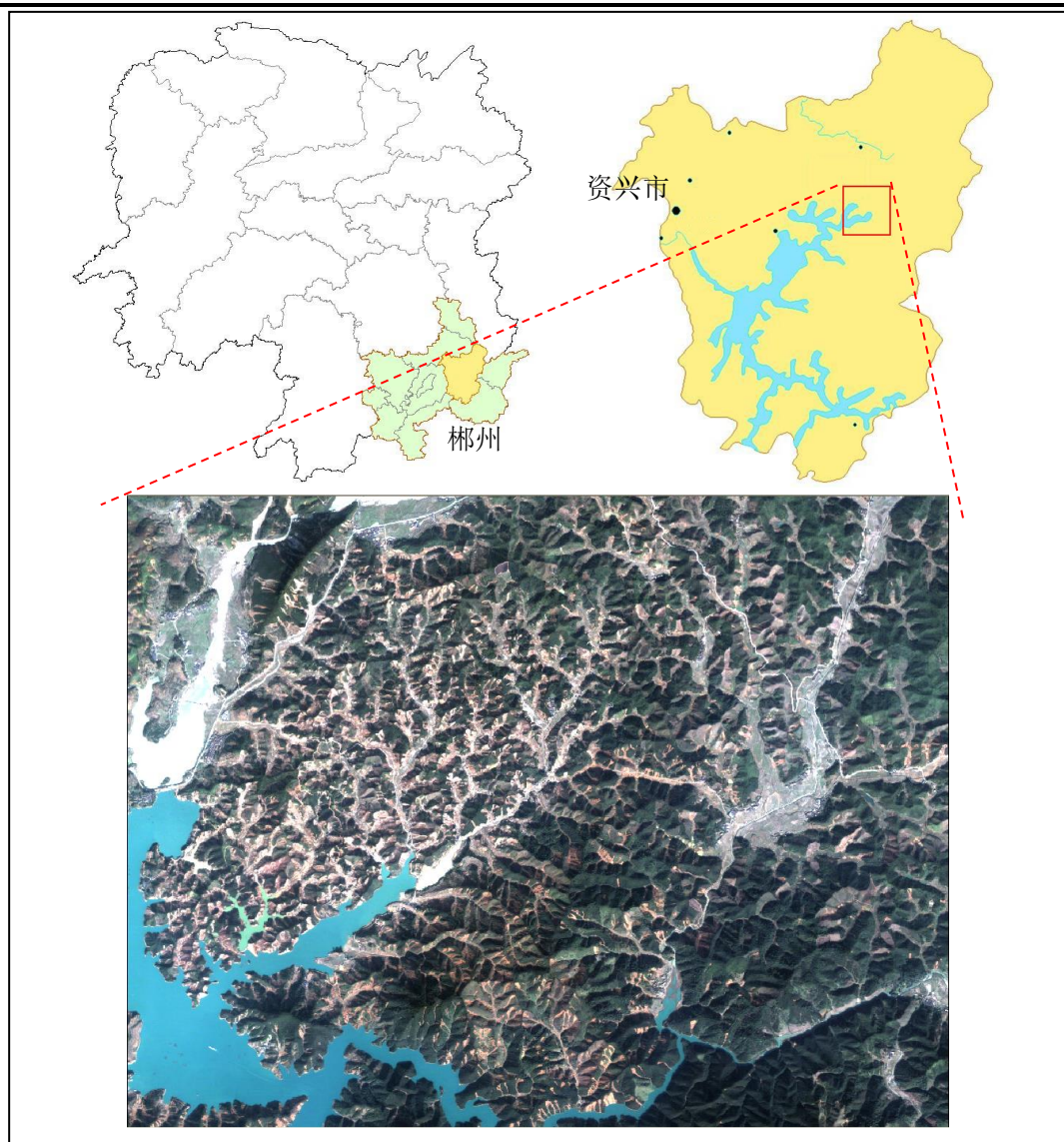


图 3-1 研究区地理位置示意图（上）及高分辨率遥感影像（下）

3.3.2 基于 SSC 的滑坡对象构建

面向对象遥感滑坡信息提取的第一个关键步骤就是滑坡影像对象构建，也即是影像分割步骤。本实验采用 eCognition 软件提供的多尺度分割技术，结合 3.2 节提出的 SSC 分割参数优化方法来实现滑坡对象构建过程。其中，分割质量的评价指标计算通过 ArcGIS 软件的 Field Calculator 功能计算得到；目标函数拟合通过 Matlab 的 Curve Fitting Tool (cftool) 工具中三次样条曲线 (cubic spline) 拟合实现。基于 SSC 的滑坡对象构建整体流程如图 3-2 所示：

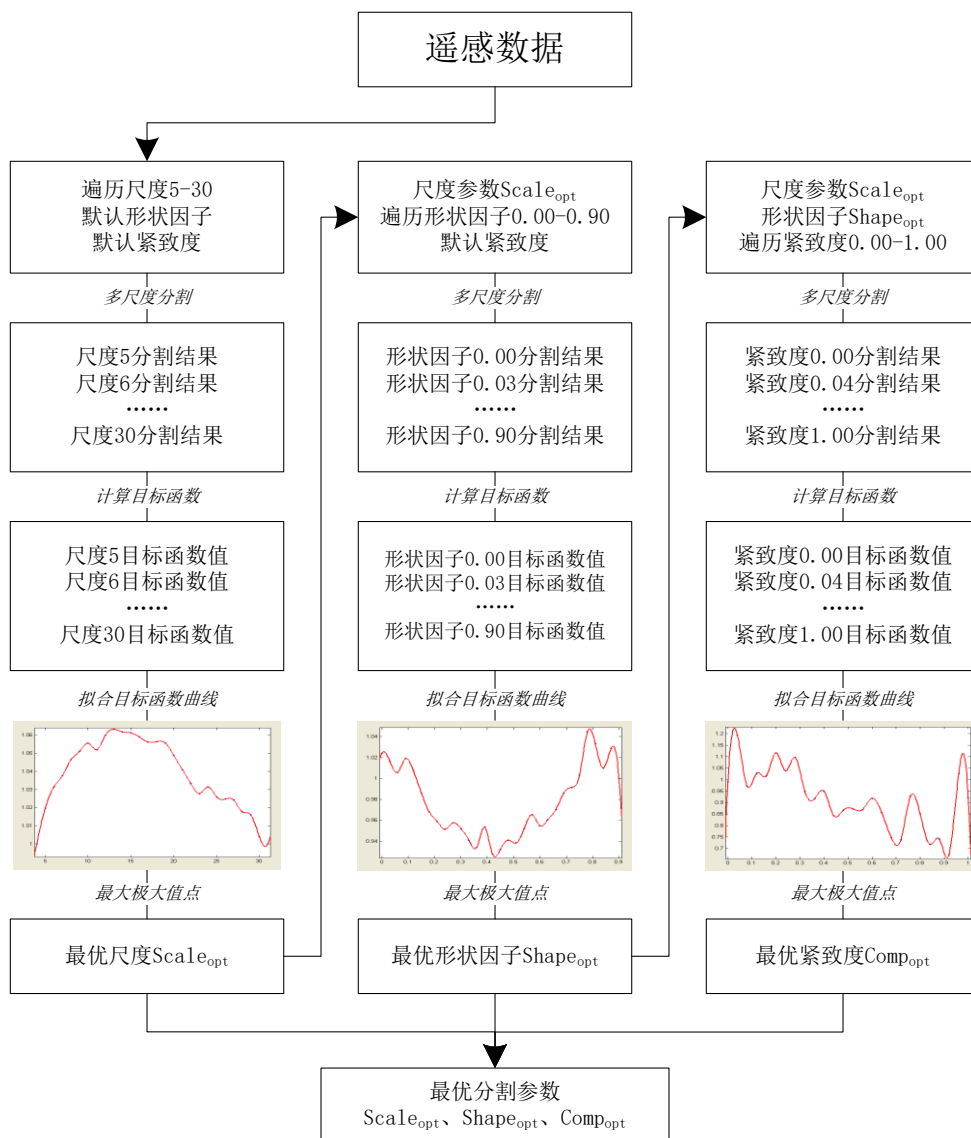


图 3-2 基于 SSC 的滑坡对象构建整体流程图

基于 SSC 的滑坡对象构建主要需要优化 3 个参数，分别是尺度参数（ $Scale_{opt}$ ）、形状因子参数（ $Shape_{opt}$ ）和紧致度参数（ $Comp_{opt}$ ）。

（1）尺度参数优化

首先，将遥感影像导入 eCognition 软件，形状因子和紧致度分别采用默认值 0.1 和 0.5，尺度范围为 5 至 30，以 1 为步长进行遍历，形成 26 个样本分割参数。然后，将 26 个样本分割参数分别代入多尺度分割函数，产生 26 个分割结果。分别将这 26 个分割结果及属性特征（包括对象亮度值、亮度方差、面积等计算分割质量评价需要用到属性特征）导出，利用 ArcGIS 软件计算各个结果的分割质量评价。最后，以尺度为横坐标，分割质量评价值为纵坐标，通过 Matlab 软件拟合得到尺度参数优化三次样条曲线（如图 3-3 所示），计算最大极大值，也即尺度参数优化值 $Scale_{opt}$ 为 12.935。

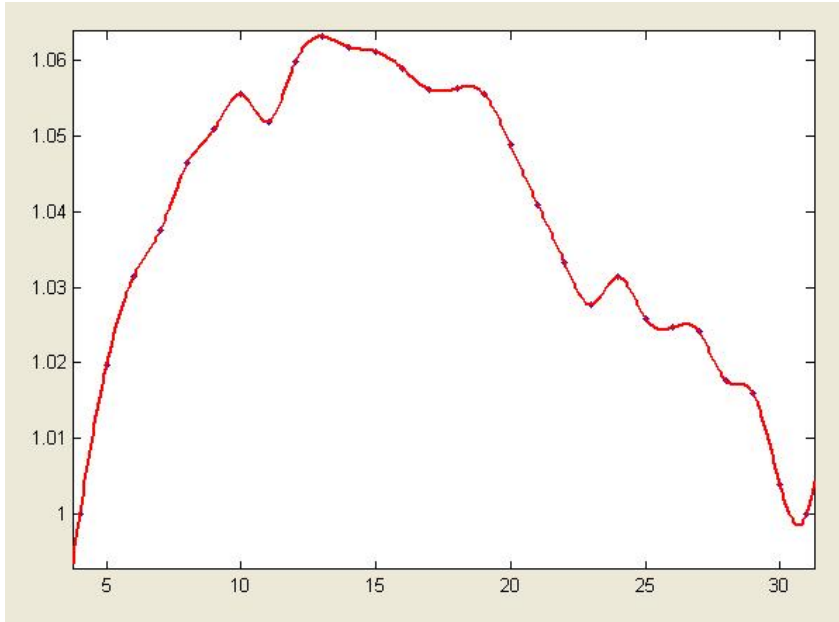


图 3-3 尺度参数优化三次样条曲线

(2) 形状因子参数优化

首先，将遥感影像导入 eCognition 软件，尺度参数设置为 12.935 ($Scale_{opt}$)，紧致度采用默认值 0.5，形状因子范围为 0 至 0.9，以 0.03 为步长进行遍历，形成 31 个样本分割参数。然后，将 31 个样本分割参数分别代入多尺度分割函数，产生 31 个分割结果。分别将这 31 个分割结果及属性特征（包括对象亮度值、亮度方差、面积等计算分割质量评价值需要用到的属性特征）导出，利用 ArcGIS 软件计算各个结果的分割质量评价值。最后，以尺度为横坐标，分割质量评价值为纵坐标，通过 Matlab 软件拟合得到形状因子参数优化三次样条曲线（如图 3-4 所示），计算最大极大值，也即尺度参数优化值 $Shape_{opt}$ 为 0.788。

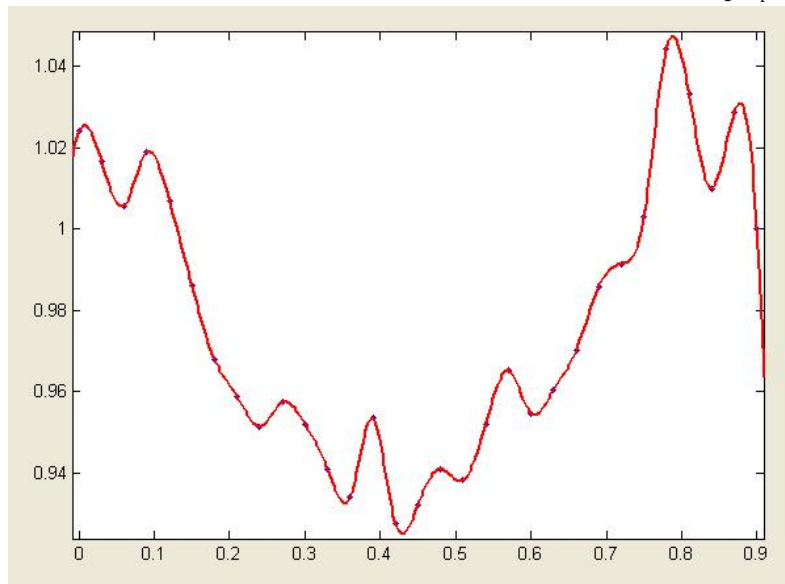


图 3-4 形状因子参数优化三次样条曲线

(3) 紧致度参数优化

首先, 将遥感影像导入 eCognition 软件, 尺度和形状因子分别采用优化值 12.935 ($Scale_{opt}$) 和 0.788 ($Shape_{opt}$), 紧致度范围为 0 至 1, 以 0.04 为步长进行遍历, 形成 26 个样本分割参数。然后, 将 26 个样本分割参数分别代入多尺度分割函数, 产生 26 个分割结果。分别将这 26 个分割结果及属性特征 (包括对象亮度值、亮度方差、面积等计算分割质量评价需要用到的属性特征) 导出, 利用 ArcGIS 软件计算各个结果的分割质量评价。最后, 以尺度为横坐标, 分割质量评价为纵坐标, 通过 Matlab 软件拟合得到紧致度参数优化三次样条曲线 (如图 3-5 所示), 计算最大极大值, 也即尺度参数优化值 $Comp_{opt}$ 为 0.028。

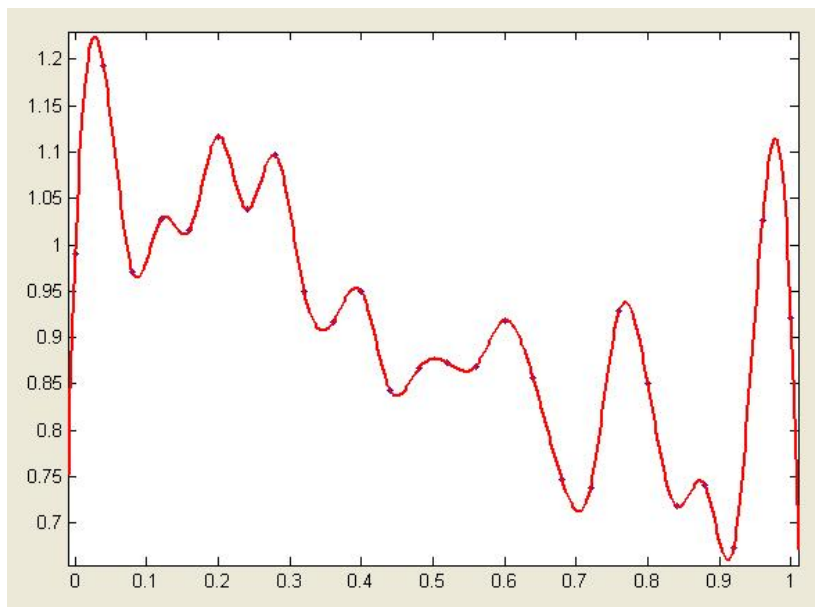


图 3-5 紧致度参数优化三次样条曲线

综合尺度、形状因子和紧致度 3 个参数的优化结果, 即可得到最优分割参数 Seg_{opt} (12.935, 0.788, 0.028)。利用最优分割参数 Seg_{opt} 得到的多尺度分割结果如图 3-6 和图 3-7:

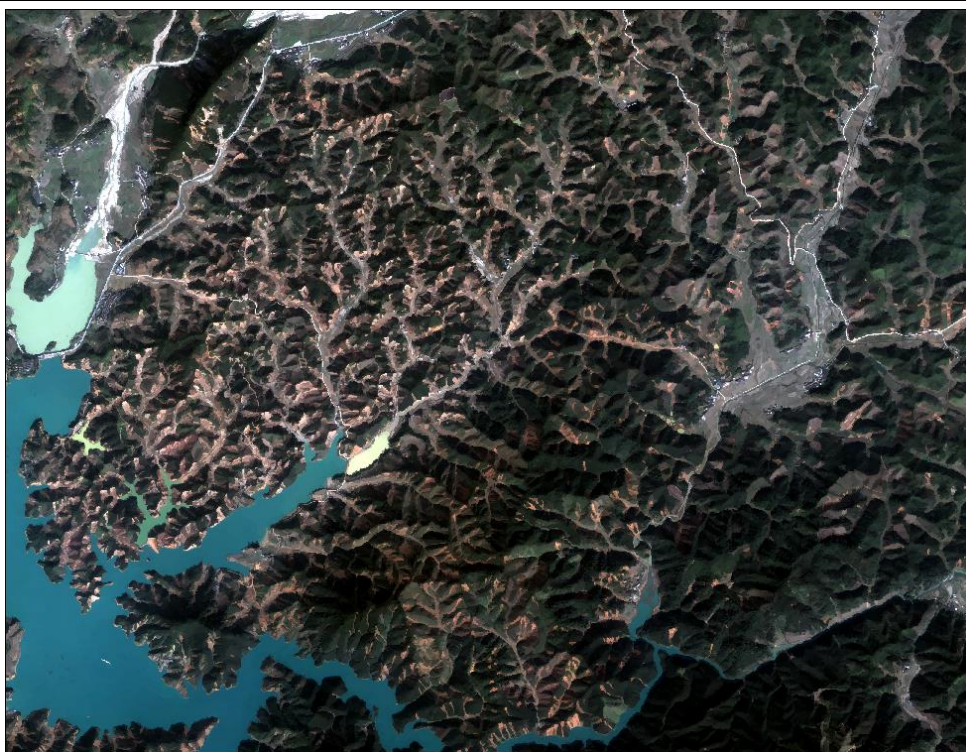


图 3-6 Seg_{opt} 多尺度分割结果（对象均值显示）

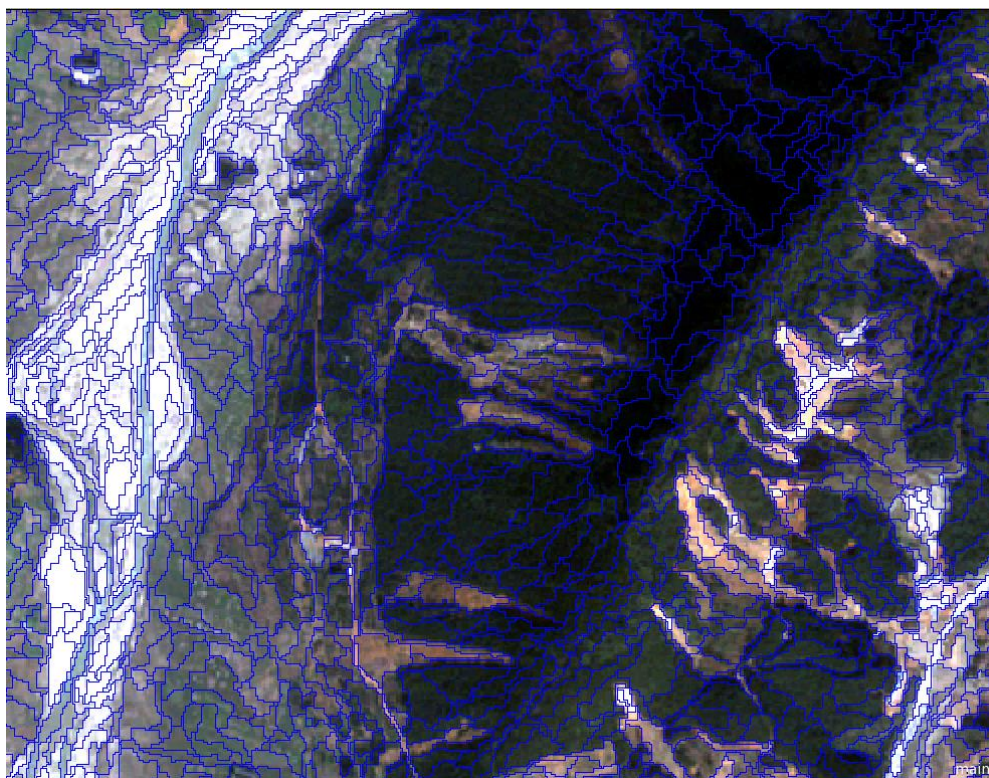











图 3-7 Seg_{opt} 多尺度分割结果（局部对象边界显示）

3.3.3 分割结果验证

将最优分割参数 Seg_{opt} (12.935, 0.788, 0.028) 多尺度分割结果, 与滑坡目视解译结果和默认分割参数 Seg_{def} (10, 0.1, 0.5) 分割结果进行对比。实验选取了 3 个滑坡地物的分割结果进行比较, 具体对比结果如表 3-1:

表 3-1 分割结果对比表

滑坡 编号	滑坡目视解译结果	默认分割结果 Seg_{def}	最优分割结果 Seg_{opt}
1			
2			
3			

从表 3-1 对比结果可以发现, 采用默认分割参数得到的分割结果中影像对象太小, 不能很好的表示目标地物, 且影像对象过多会导致信息提取效率降低。基于 SSC 的滑坡对象构建方法, 在不具备先验知识的前提下, 可以计算出一个分割参数, 使得分割后产生的滑坡对象与实际目标大小最为接近。由此可见在面向对象遥感滑坡信息提取中, 基于 SSC 的滑坡对象构建方法对于如何设置分割参数具有重要的指导意义和应用价值。

§3.4 本章小结

本章主要研究基于 SSC 的滑坡对象构建方法。首先, 介绍多尺度分割的概念、算法和

技术流程，然后，提出 SSC (Scale-Shape-Compactness) 多尺度分割参数优化方法，最后，进行基于 SSC 的滑坡对象构建实验。为使得到的分割结果能够有效地确保滑坡地物边界提取的准确性，SSC 多尺度分割参数优化方法，利用对象内方差和莫兰 I 指数组成的分割质量评价函数，从而针对多尺度分割中的尺度、形状因子和紧致度参数进行优化。在面向对象遥感滑坡信息提取中，该方法对于如何设置多尺度分割参数具有重要的指导意义和应用价值。

第四章 随机森林滑坡特征选择方法

随机森林常被用于解决回归问题、分类问题以及特征选择问题，是一种灵活的机器学习方法。由于随机森林方法在特征选择方面表现出明显优势，因而将其应用于面向对象滑坡信息提取的特征选择中。本章首先介绍随机森林的基本原理，然后介绍面向对象的随机森林滑坡特征选择方法，之后在此基础上进行滑坡特征选择实验，最后对随机森林滑坡特征选择方法进行总结。

§4.1 随机森林的基本原理

随机森林是由 Leo Breiman 和 Adele Cutler 于 2001 年共同发展出的一种有效的机器学习方法。作为一种集成分类器，随机森林由多个决策树共同组成。

常见的决策树模型有：ID3、CART、C4.5 和 C5.0 等。在决策树模型中，每一个决策节点都对应的用一个非叶子节点来表示，利用某个特征的特定规则对其进行约束和划分，直到符合终止条件。连续性数据和离散型数据都可以用决策树来处理。决策树最上面的节点是它的根节点，所有的训练集都包含在根节点中。如果某一类样本不被节点所包含，那么要继续对其进行分割，直到所有节点都被包含为止。由于在对节点进行分割时，我们需要用所有的特征去尝试，因此每一个特征的划分方式都是多种多样的，在遍历了所有特征的分割方式之后，选择最好的方式，对该节点进行最优分割。在决策树中，我们通常用信息增益（Information Gain）和基尼指数（Gini Index）^[63]来衡量在某种分割方式下特征的分割效果。

当采取信息增益来衡量分割效果时，用 $I(S)$ 来表示分类的期望信息：

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m P_i \log_2(P_i) \quad (4-1)$$

公式 4-1 中， S 表示的是样本集合， S 中含有样本的类别数用 m 来表示，将 S 的第 i 个划分表示为 s_i ，而第 i 类样本的概率则用 P_i 表示， $P_i = |s_i| / |S|$ 。样品越有序，则 $I(S)$ 的值越小。

为了验证特征的分割效果，假设用 A 特征来分割某一节点，然后计算其划分节点的熵，并将熵表示为 $E(A)$ ：

$$E(A) = \sum_{j=1}^r (s_{1j} + \dots + s_{mj}) / s \cdot I(s_{1j}, \dots, s_{mj}) \quad (4-2)$$

在上述公式中, 将分割节点的特征用 A 来表示, 假设用 V 来表示 A 的不同取值的个数, 也就是说 S 可以被划分成 V 个不同的子集, 在子集 s_j 中含有的第 i 类样品的个数用 s_{ij} 表示。在计算了期望值和熵值后, 用特征 A 分割节点的信息增益 $Gain(A)$ 也可以随之计算出:

$$Gain(A) = I(s_1, \dots, s_m) - E(A) \quad (4-3)$$

在用基尼指数来衡量分割效果时, 假设 T 集合中包含的不同类别记录的样品的个数为 N , 那么可以得到 $Gini(T)$ 如下:

$$Gini(T) = 1 - \sum_{j=1}^N P_j^2 \quad (4-4)$$

在公式 4-4 中, 将第 j 类样品出现的概率表示为 P_j 。假如经过某次分割后, T 集合被分成 m 个部分, 用 N_1, N_2, \dots, N_m 分别表示各个部分。所以可以得到经此次分割后的 $Gini_{split}(T)$ 为:

$$Gini_{split}(T) = \frac{N_1}{N} Gini(T_1) + \dots + \frac{N_m}{N} Gini(T_m) \quad (4-5)$$

随机森林是基于决策树的, 由于随机森林是修改了决策树候选分割特征集, 引入了集成思想, 所以其性能相对于决策树而言较高。随机森林的随机性体现在两个方面: 1) 训练集抽取的随机性; 2) 节点候选分割特征集合的随机性。集成方法的性能很大程度上取决于基分类器之间的差异性, 正是因为构成随机森林的决策树的随机性保证了随机森林的多样性, 使得其被引入多个研究领域。

§4.2 基于对象的随机森林滑坡特征选择

通过一系列的规则对特征进行排名, 可以得到特征重要性的相对关系。由上节内容不难得出, 分割特征是在投票环节得票最多的特征, 这也是特征的作用在划分节点中的体现。特征和类标之间的相关程度决定了特征的重要程度。特征排名的实现, 是靠随机森林对特征重要性进行衡量来得到的, 常用的重要性度量主要有以下几种:

(1) 分割特征次数度量

通常情况下, 我们需要统计特征作为分割特征的次数, 出现的频率越高说明其越重要。显然, 如此简单的来衡量一个指标的重要性是不合理的, 而且随机森林其本身的随机性也会对评价的准确与否产生影响。如果对特征的得分做差, 那么其差异必然是自然数, 而特征间的差异不可能如此整齐, 因此特征之间的真实差异也得不到很好的体现。在对节点选择候选特征时, 每个特征被选取的概率都是相等的, 所以会出现以下几种情况:

- 1) 在等概率的情况下, 一些重要的特征被抽中的机会可能很少, 有时甚至会出现没有抽中的极端情况, 那么会导致得分很低, 但此时得分低并不能说明该特征就不重要;
- 2) 有些得分较高的中上等特征会掩盖了真实的重要特征;
- 3) 由于随机性的存在, 噪音特征可能会受益, 既能凸显出其重要性又不会使其变成重要特征, 这样就维持了随机森林构建决策树的策略。

由于这种方法优缺点明显, 虽然简单, 但是不够周密, 所以应用范围有一定的限制, 接下来介绍的基尼重要性要比该方法细腻周密些。

(2) 基尼重要性度量

在决策树被构建的过程中, 节点的分割效果采用基尼系数 (Gini Index) 来进行度量, 而特征的重要性程度则通过计算基尼重要性 (Gini Importance)^[64]来实现。为避免采用统计分割次数那样粗粒度的重要性度量方法, 在利用随机森林度量特征重要性的时候, 我们引入特征分割节点对应的基尼系数。基尼系数可以反映决策节点和分割方式之间的相对关系, 若特征分割效果好, 则基尼系数小, 反之亦然。特征的最终得分是通过将决策节点得到的基尼系数进行线性累加或加权线性累加而计算得到, 这个最终得分即是基尼重要性 (Gini Importance)。由于这种得分可以精确到小数点后很多位, 因此基尼重要性度量是一种细粒度的重要性度量方法。

(3) 排列重要性度量

特征排列的重要性 (permutation Importance)^[65]可以通过计算重新排列特征在袋外数据 (OOB, out-of-bag) 样本上的取值而获得。在众多的随机森林特征选择方法中, 该方法是使用最多的。在进行特征的排列重要性得分计算时, 是以每棵决策树作为出发点的。要想计算某棵决策树对一个特征的重要性的度量结果, 首先必须保证其它特征不发生变化, 然后对该树的袋外数据样本在这个特征上的取值分布进行重新排列。再利用此树对“新”的 OOB 样本类标做预测。将 OOB 样本与类标对应关系打乱前与打乱后所预测的正确率做差, 得到的差值即是该决策树的特征重要性度量结果。这种差值分为三种情况:

- 1) 差值为正值。若该特征是一个重要特征且在不同样本间可明显区分, 则经过随机重排以后, 其区分度会下降, 得分为正值;
- 2) 差值为负值。假如该特征对样本而言是非重要特征且混乱分布于不同样本间, 则经过随机重排以后, 其混乱程度可能会降低, 得分为负值;
- 3) 差值为零。若该特征对样本而言无用, 那么样本在此特征上必然与类标无关, 此时不管进行怎样的重排, 得到的结果都是不变的, 得分为零。

面向对象遥感滑坡信息提取中的影像分割过程将构建出许多影像对象, 从 2.1.2 节介绍的内容中很容易看出, 影像对象具有丰富的特征信息, 这些特征较单个像元特征更接近地理实体的信息。若将全部特征应用到滑坡信息提取中难免会造成“Hughes”现象^[57]。因此, 特征选择是面向对象遥感滑坡信息提取的重要步骤之一。依据之前对随机森林的介绍, 可以发现, 随机森林适合处理小样本大数据量, 在特征选择方面表现出了精度高、速度快、稳定性好的特点。因此, 我们将面向对象影像分析技术与随机森林特征选择方法相结合, 通过 R 语言^[66]编写的程序, 读取样本影像对象的特征信息, 同时调用随机森林软件包^[67]对候选特征进行运算, 以得遥感滑坡信息提取中各类地物的重要特征, 并按重要性进行排

序。接下来，将进行随机森林滑坡特征选择实验，以验证该方法的适用性。

§4.3 随机森林滑坡特征选择实验

4.3.1 实验数据及预处理

随机森林滑坡特征选择实验数据使用的实验数据来自荷兰屯特大学（University of Twente）地理信息科学与地球观测学院（Faculty of Geo-Information Science and Earth Observation, ITC）Martha 等在网上共享的示例数据（<http://www.itc.nl/OOA-group>），包括多光谱数据和 DEM 数据：

- 1）多光谱数据是由搭载在 IRS-P6（Resourcesat-1）卫星上线性成像自扫描（LISS-IV）传感器于 2004 年 4 月 16 日获取^[68]，包括近红外（0.76-0.86μm）、红（0.62-0.68μm）、绿（0.52-0.59μm）三个波段，星下点几何分辨率为 5.8m，主要用来参与影像的分割运算，以及提取光谱诊断性特征；
- 2）DEM 数据由 IRS-P5（Cartosat-1）卫星于 2006 年 4 月 6 日获取的立体像对通过数字摄影测量的方法制作而成^[69]，空间分辨率为 10m，垂直方向上均方根误差（RMSE）为 2.31m，主要用于产生坡度、坡向等 DEM 衍生数据。

DEM 的衍生数据包括：坡度、坡向、山体阴影、地表曲率、坡面曲率、河网等，我们利用 GIS 软件对 DEM 数据进行地形分析即可得到（图 4-1）。其中需要注意的是山体阴影的计算需要提供多光谱影像获取时的太阳方位，这些可以在影像的头文件中得到，本实验数据的太阳高度角为 64°，太阳方位角为 135°。在研究区缺少河网矢量的情况下，通过对 DEM 进行水文分析，可以得到河网信息。实验中水系的提取 1）先通过 DEM 计算出水流方向，2）之后由水流方向计算汇流累积量，3）然后由水流方向和汇流累积量用 Strahler 方法计算出河网等级，4）最后提取河流等级高（≥5）的主河道，形成水系矢量图。

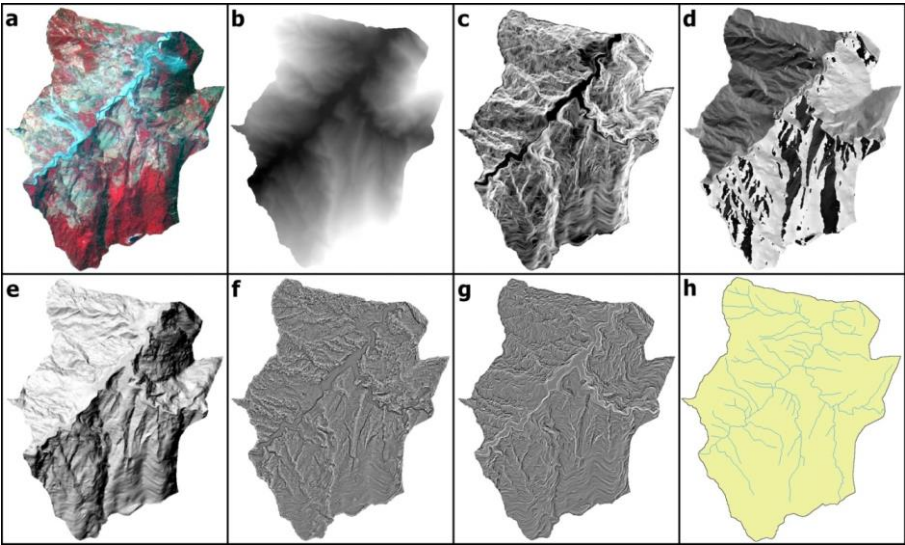


图 4-1 多源实验数据

(a)多光谱影像, (b)DEM, (c)坡度, (d)坡向, (e)山体阴影, (f)地表曲率, (g)坡面曲率, (h)水系

4.3.2 选择候选影像对象特征

本文的 2.1.2 节中已经对面向对象信息提取中使用的特征进行了详细介绍，本实验从中选取了覆盖自定义特征、波段均值、波段标准差、与相邻对象的差异、几何范围、几何形状、纹理同质性灰度共生矩阵和纹理均值灰度共生矩阵的 76 个候选特征（如表 4-1）进行研究。

表 4-1 影像对象候选特征

特征类型		特征名	说明
自定义特征		Asp-MainDir	坡向与影像对象主轴方向的夹角
		Relief	相对高差
图 层 值	均值	Mean Aspect	坡向均值
		Mean Brightness	亮度均值
		Mean DEM	数字高程模型均值
		Mean Layer1	绿色波段均值
		Mean Layer2	红色波段均值
		Mean Layer3	近红外波段均值
		Mean Max. diff.	最大差异均值
		Mean curvature	地表曲率均值
		Mean flow_dir	水流方向均值
		Mean hill_shade	山体阴影均值
		Mean ndvi	植被指数均值
		Mean prof_curv	坡面曲率均值
		Mean slope	坡度均值
	标准差	Standard deviation Aspect	坡向标准差
		Standard deviation DEM	数字高程模型标准差
		Standard deviation Layer1	绿色波段标准差
		Standard deviation Layer2	红色波段标准差
		Standard deviation Layer3	近红外波段标准差
		Standard deviation curvature	地表曲率标准差
		Standard deviation flow_dir	水流方向标准差
		Standard deviation hill_shade	山体阴影标准差
		Standard deviation ndvi	植被指数标准差
		Standard deviation prof_curv	坡面曲率标准差
		Standard deviation slope	坡度标准差
	与邻接对象的差异	Mean Diff. to neighbors Aspect	坡向与邻接对象的差异
		Mean Diff. to neighbors DEM	数字高程模型与邻接对象的差异
		Mean Diff. to neighbors Layer1	绿色波段与邻接对象的差异
		Mean Diff. to neighbors Layer2	红色波段与邻接对象的差异

几何特征		Mean Diff. to neighbors Layer3	近红外波段与邻接对象的差异
		Mean Diff. to neighbors curvature	地表曲率与邻接对象的差异
		Mean Diff. to neighbors flow_dir	水流方向与邻接对象的差异
		Mean Diff. to neighbors hill_shade	山体阴影与邻接对象的差异
		Mean Diff. to neighbors ndvi	植被指数与邻接对象的差异
		Mean Diff. to neighbors prof_curv	坡面曲率与邻接对象的差异
		Mean Diff. to neighbors slope	坡度与邻接对象的差异
	范围	Area	面积
		Border length	周长
		Length	对象长轴长度
		Length/Width	长宽比
		Number of pixels	像素个数
		Width	对象短轴长度
	形状	Asymmetry	不对称性
		Border index	边界指数
		Compactness	紧致度
		Density	密度
		Elliptic Fit	椭圆拟合度
		Main Direction	主方向，即对象长轴的方向
		Radius of largest enclosed ellipse	最大的封闭式椭圆半径
		Radius of smallest enclosing ellipse	最小包围椭圆半径
		Rectangular Fit	矩形拟合度
		Roundness	圆度
		Shape Index	形状指数
纹理特征	同质性灰度共生矩阵	GLCM Homogeneity (All dir) Aspect	坡向同质性灰度共生矩阵
		GLCM Homogeneity (All dir) DEM	数字高程模型同质性灰度共生矩阵
		GLCM Homogeneity (All dir) Layer1	绿色波段同质性灰度共生矩阵
		GLCM Homogeneity (All dir) Layer2	红色波段同质性灰度共生矩阵
		GLCM Homogeneity (All dir) Layer3	近红外波段同质性灰度共生矩阵
		GLCM Homogeneity (All dir) curvature	地表曲率同质性灰度共生矩阵
		GLCM Homogeneity (All dir) flow_dir	水流方向同质性灰度共生矩阵
		GLCM Homogeneity (All dir) hill_shade	山体阴影同质性灰度共生矩阵
		GLCM Homogeneity (All dir) ndvi	植被指数同质性灰度共生矩阵
		GLCM Homogeneity (All dir) prof_curv	坡面曲率同质性灰度共生矩阵
		GLCM Homogeneity (All dir) slope	坡度同质性灰度共生矩阵
	均值灰度共生矩阵	GLCM Mean (All dir) Aspect	坡向均值灰度共生矩阵
		GLCM Mean (All dir) DEM	数字高程模型均值灰度共生矩阵
		GLCM Mean (All dir) Layer1	绿色波段均值灰度共生矩阵

	GLCM Mean (All dir) Layer2	红色波段均值灰度共生矩阵
	GLCM Mean (All dir) Layer3	近红外波段均值灰度共生矩阵
	GLCM Mean (All dir) curvature	地表曲率均值灰度共生矩阵
	GLCM Mean (All dir) flow_dir	水流方向均值灰度共生矩阵
	GLCM Mean (All dir) hill_shade	山体阴影均值灰度共生矩阵
	GLCM Mean (All dir) ndvi	植被指数均值灰度共生矩阵
	GLCM Mean (All dir) prof_curv	坡面曲率均值灰度共生矩阵
	GLCM Mean (All dir) slope	坡度均值灰度共生矩阵

4.3.3 特征选择实验流程

实验采用 eCognition 作为影像的多尺度分割工具。通过参考研究区的调查数据以及目视解译结果，选取滑坡信息提取过程中涉及到的各类地物样本对象，并将样本对象和其对应的候选特征信息通过 eCognition 中导出。然后用 R 语言编写的随机森林特征选择程序对导出的数据进行计算。最后得到各类地物的随机森林特征选择结果以及各个特征的贡献度。具体的实验流程包括：

- 1) 数据预处理。该过程在 4.3.1 节已经做过详细介绍；
- 2) 影像对象构建。运用第三章介绍的 SSC 多尺度分割参数优化方法得到优化的分割参数 Seg_{opt} (10, 0.1, 0.5)，利用优化参数对影像进行多尺度分割，构建影像对象；
- 3) 选取样本对象。参考 Martha 等人在研究区的信息提取结果，参考研究区的调查数据以及目视解译结果，选取滑坡信息提取过程中涉及到的各类地物样本对象，包括阴影、河流、湖泊、滩涂、农用地、建筑物、树木、岩区、陡崖、道路等地物，同时将这些样本对象和其对应的候选特征信息通过 eCognition 中导出；
- 4) 随机森林特征选择。用 R 语言编写随机森林特征选择程序，对导出的样本特征数据进行计算，得到各类地物的特征选择结果以及各个特征的贡献度。

4.3.4 实验结果与分析

(1) 阴影

阴影地物特征选择结果如图 4-2 所示，实验共选出了 34 个特征，其中山体阴影均值特征贡献度最大，达到 13.32；贡献度最小的特征是近红外波段的标准差，仅为 2.73。

所选特征集贡献度均值为 5.34，大于均值的特征包括：1.山体阴影均值；2.山体阴影均值灰度共生矩阵；3.最大差异均值；4.植被指数均值；5.近红外波段均值灰度共生矩阵；6.近红外波段均值；7.DEM 同质性灰度共生矩阵；8.亮度均值；9.坡度均值灰度共生矩阵；10.坡度均值；11.坡向均值灰度共生矩阵；12.近红外波段同质性灰度共生矩阵；13.坡向均值。

在滑坡信息提取中，根据专家知识设定的阴影地物提取规则为：山体阴影≤92；植被指数≥0.1；亮度≤45。可以发现该规则使用的特征与随机森林选取的特征有着高度的一致性，且贡献度排名分别为第一、第四和第八。另外，最大差异均值、近红外波段、DEM、坡度和坡向等特征对于提取阴影也具有一定指导作用和启发意义。

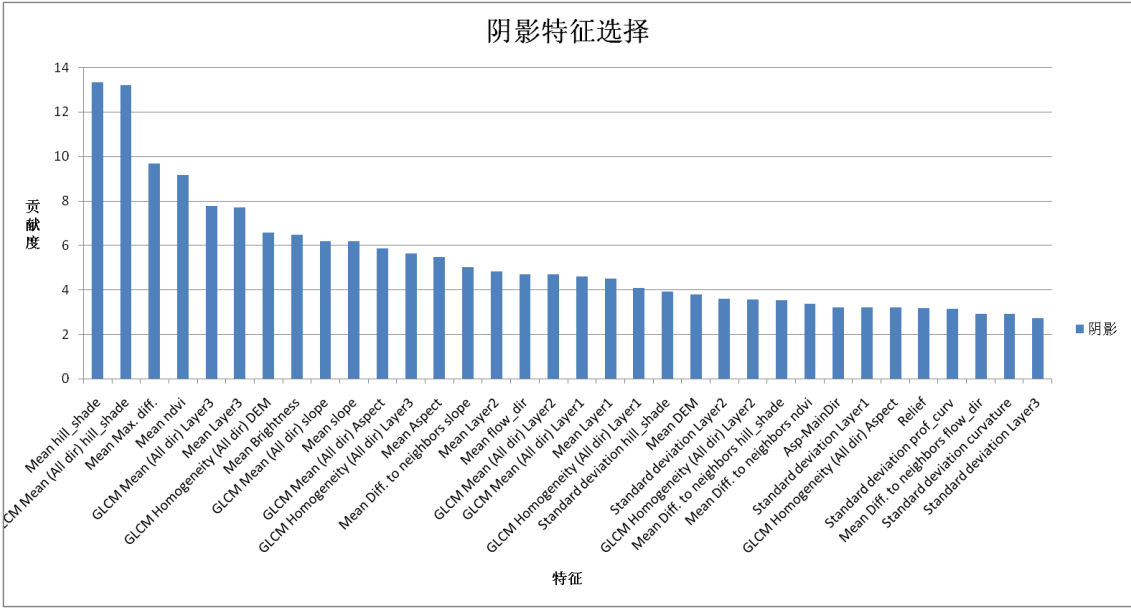


图 4-2 阴影特征选择结果

(2) 河流

河流地物特征选择结果如图 4-3 所示，实验共选出了 37 个特征，其中最大差异均值贡献度最大，达到 12.07；贡献度最小的特征是边界指数，仅为 3.04。

所选特征集贡献度均值为 5.27，大于均值的特征包括：1.最大差异均值；2.植被指数均值；3.坡度均值；4.坡度均值灰度共生矩阵；5.近红外波段均值灰度共生矩阵；6.相对高差；7.近红外波段均值；8.DEM 标准差；9.近红外波段与邻接对象的差异；10.近红外波段同质性灰度共生矩阵；11.红色波段与邻接对象的差异；12.亮度均值。

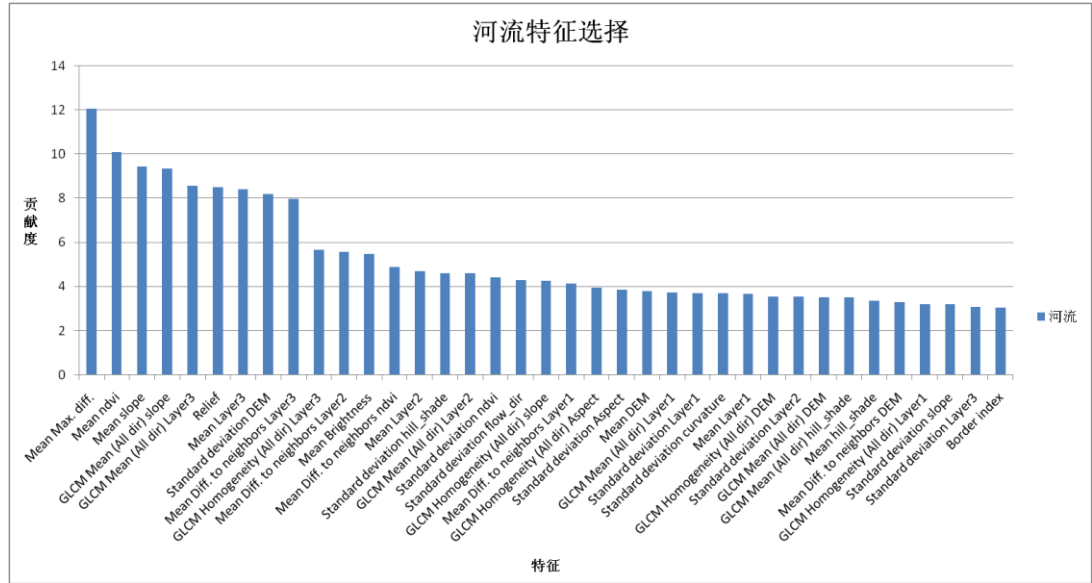


图 4-3 河流特征选择结果

在滑坡信息提取中，根据专家知识设定的河流地物提取规则为：河网等级 ≥ 5 。河网等

级特征是由 DEM 通过数字地形分析计算得到的专题特征，因为该特征在 eCognition 中为变量特征，未列入候选特征，所以河流的特征集中并未出现河网等级特征。但从另一方面考虑，河网等级是通过 DEM 计算得到，而 DEM 在河流的特征集中排名第八。因此在河流提取中，专家经验特征与随机森林选取的特征任然保持有一定的一致性。另外，最大差异均值、植被指数、坡度、近红外波段、相对高差、DEM、红色波段和亮度等特征对于提取河流也具有一定指导作用和启发意义。

(3) 湖泊

湖泊地物特征选择结果如图 4-4 所示，实验共选出了 14 个特征，其中近红外波段均值灰度共生矩阵特征贡献度最大，达到 4.88；贡献度最小的特征是植被指数均值，仅为 2.34。

所选特征集贡献度均值为 3.30，大于均值的特征包括：1.近红外波段均值灰度共生矩阵；2.近红外波段均值；3.亮度均值；4.DEM 标准差；5.相对高差。

在滑坡信息提取中，根据专家知识设定的湖泊地物提取规则为：近红外波段 ≤ 55 ；坡度 ≤ 5 。其中，近红外波段贡献度排名靠前，坡度排名居中。因此在湖泊提取中，专家经验特征与随机森林选取的特征保持有一定的一致性。另外，近红外波段、亮度、DEM 和相对高差等特征对于提取湖泊也具有一定指导作用和启发意义。

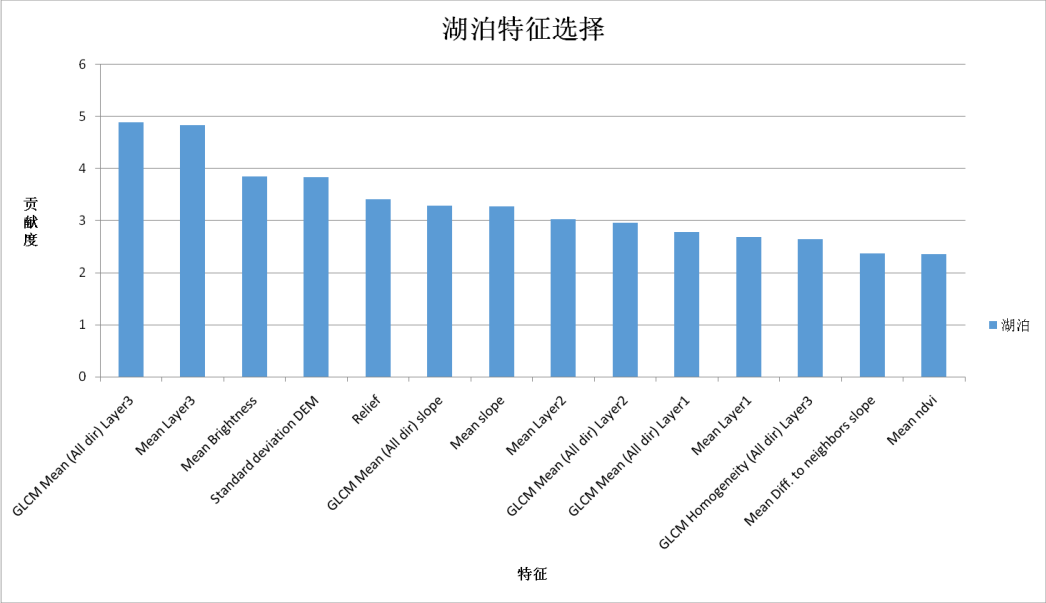


图 4-4 湖泊特征选择结果

(4) 滩涂

滩涂地物特征选择结果如图 4-5 所示，实验共选出了 43 个特征，其中坡度均值灰度共生矩阵特征贡献度最大，达到 12.60；贡献度最小的特征是山体阴影同质性灰度共生矩阵，仅为 2.92。

所选特征集贡献度均值为 5.86，大于均值的特征包括：1.坡度均值灰度共生矩阵；2.坡度均值；3.DEM 标准差；4.相对高差；5.DEM 均值；6.DEM 均值灰度共生矩阵；7.植被指数均值；8.最大差异均值；9.近红外波段均值灰度共生矩阵；10.DEM 同质性灰度共生矩阵；11.近红外波段均值；12.近红外波段同质性灰度共生矩阵；13.红色波段均值；14.红色

波段均值灰度共生矩阵；15.亮度均值；16.坡度同质性灰度共生矩阵；17.植被指数标准差。

在滑坡信息提取中，根据专家知识设定的滩涂物提取规则为：坡度 ≤ 20 ；相对高差 ≤ 30 ；亮度 ≥ 65 ；与河流的距离 $\leq 100\text{m}$ 。其中，坡度和相对高差贡献度排名靠前，分别为第二和第四，亮度排名居中，仅为第十五。“与河流的距离”属于临近对象相关特征，未列入候选特征，但是河流特征选择集当中的大部分特征在滩涂特征选择集中，说明“与河流的距离”也有一定的特征贡献度。因此在滩涂提取中，专家经验特征与随机森林选取的特征保持有一定的一致性。另外，DEM、植被指数、最大差异均值、近红外波段和红色波段等特征对于提取滩涂也具有一定指导作用和启发意义。

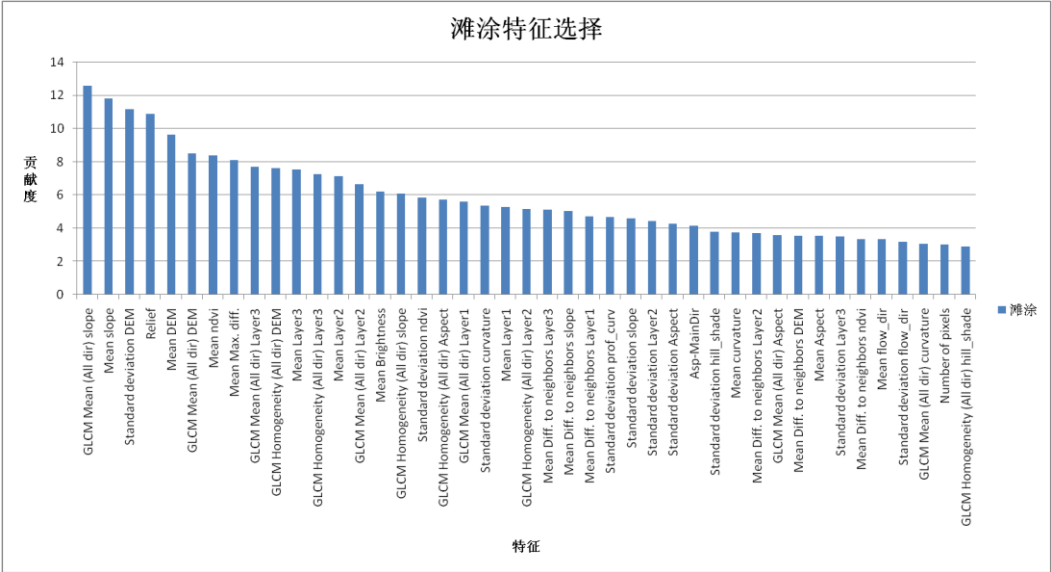


图 4-5 滩涂特征选择结果

(5) 建筑区

建筑区地物特征选择结果如图 4-6 所示，实验共选出了 14 个特征，其中坡度均值贡献度最大，达到 6.96；贡献度最小的特征是山体阴影均值灰度共生矩阵，仅为 3.12。

所选特征集贡献度均值为 4.35，大于均值的特征包括：1.坡度均值；2.坡度均值灰度共生矩阵；3.水流方向标准差；4.坡向均值灰度共生矩阵；5.相对高差。

在滑坡信息提取中，根据专家知识设定的建筑物提取规则为：坡度 ≤ 12 ；红色波段同质性灰度共生矩阵: [0.15,0.2]。其中，坡度贡献度排名位于第一位。“红色波段同质性灰度共生矩阵”并未被随机森林特征选择所选出。因此在建筑区提取中，专家经验特征与随机森林选取的特征有些许的不一致性。这可能是由于选取的建筑区样本不够典型，或者“红色波段同质性灰度共生矩阵”并不是建筑区诊断性特征。另外，水流方向、坡向和相对高差等特征对于提取建筑区也具有一定指导作用和启发意义。

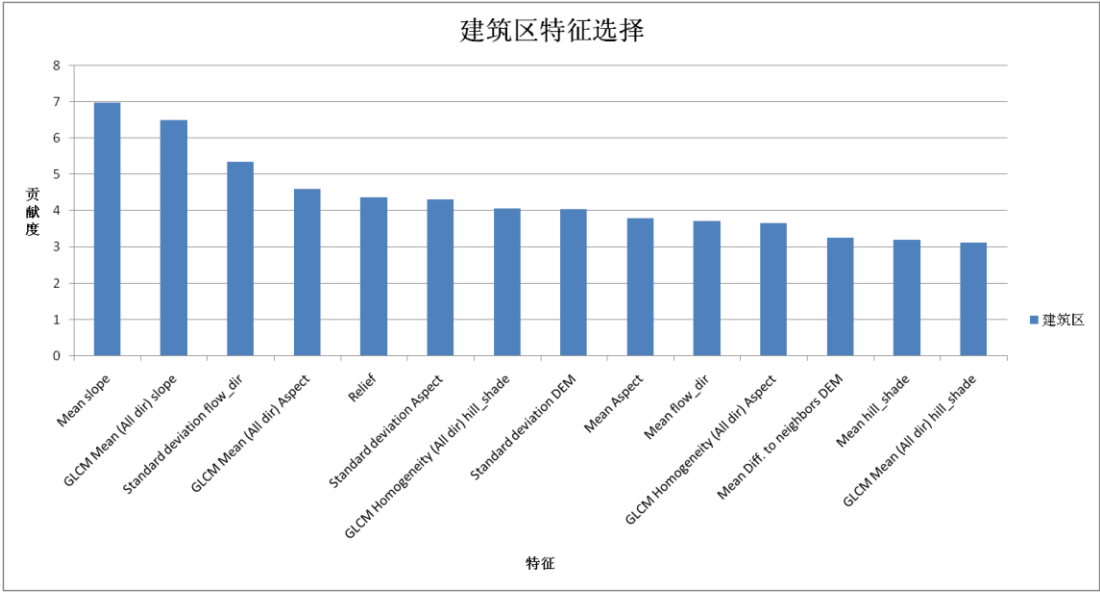


图 4-6 建筑区特征选择结果

（6）农用地

农用地特征选择结果如图 4-7 所示，实验共选出了 56 个特征，其中植被指数贡献度最大，达到 24.38；贡献度最小的特征是密度特征，仅为 3.47。

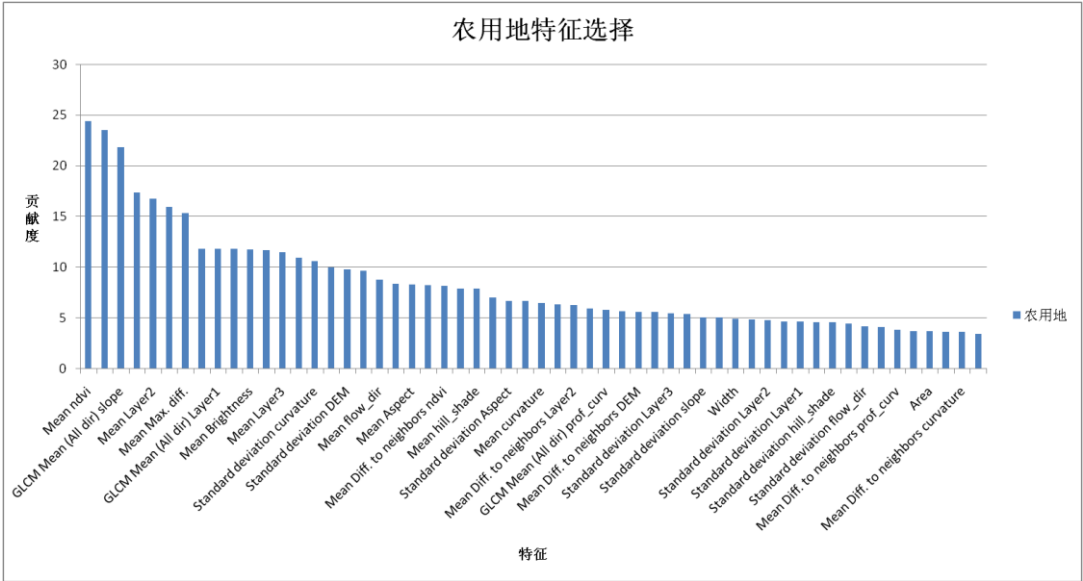


图 4-7 农用地特征选择结果

所选特征集贡献度均值为 8.42，大于均值的特征包括：1.植被指数均值；2.坡度均值；3.坡度均值灰度共生矩阵；4.红色波段均值灰度共生矩阵；5.红色波段均值；6.DEM 同质性灰度共生矩阵；7.最大差异均值；8.绿色波段均值；9.绿色波段均值灰度共生矩阵；10.近红外波段均值灰度共生矩阵；11.亮度均值；12.DEM 均值灰度共生矩阵；13.近红外波段均值；14.DEM 均值；15.地表曲率标准差；16.相对高差；17.DEM 标准差；18.坡面曲率标准差；

19.水流方向均值。

在滑坡信息提取中，根据专家知识设定的农用地提取规则为：植被指数 ≥ 0.095 ；坡度 ≤ 12 ；红色波段均值灰度共生矩阵:[60,90]。其中，植被指数、坡度和红色波段均值灰度共生矩阵的贡献度排名均靠前，分别为第一、第二和第四。因此在农用地提取中，专家经验特征与随机森林选取的特征保持有高度的一致性。另外，DEM、最大差异均值、绿色波段、近红外波段、亮度、地表曲率和水流方向等特征对于提取农用地也具有一定指导作用和启发意义。

(7) 岩区

岩区特征选择结果如图 4-8 所示，实验共选出了 54 个特征，其中坡度贡献度最大，达到 25.53；贡献度最小的特征是水流方向与邻接对象的差异，仅为 2.86。

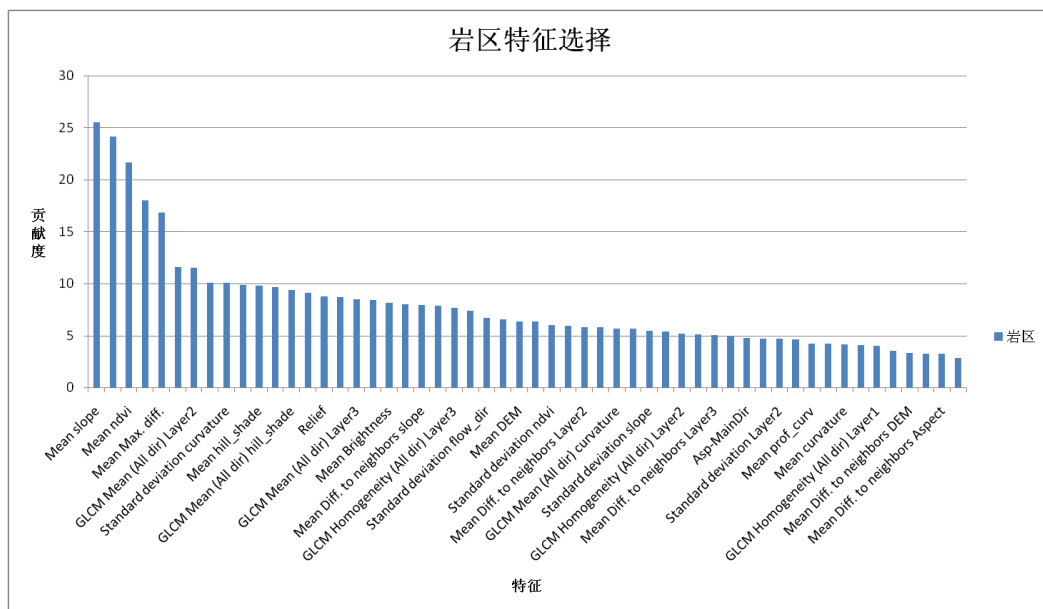


图 4-8 岩区特征选择结果

所选特征集贡献度均值为 7.89，大于均值的特征包括：1.坡度均值；2.坡度均值灰度共生矩阵；3.植被指数均值；4.DEM 同质性灰度共生矩阵；5.最大差异均值；6.红色波段均值；7.红色波段均值灰度共生矩阵；8.绿色波段均值灰度共生矩阵；9.地表曲率标准差；10.绿色波段均值；11.山体阴影均值；12.坡面曲率标准差；13.山体阴影均值灰度共生矩阵；14.植被指数与邻接对象的差异；15.相对高差；16.DEM 标准差；17.近红外波段均值灰度共生矩阵；18.近红外波段均值；19.亮度均值；20.水流方向均值；21.坡度与邻接对象的差异；22.坡向同质性灰度共生矩阵。

在滑坡信息提取中，根据专家知识设定的岩区提取规则为：坡度:[30,45]；植被指数 ≥ 0.12 ；亮度 ≤ 95 。其中，坡度和植被指数的贡献度排名靠前，分别为第一和第三，亮度排名居中，仅为第十九位。因此在农用地提取中，专家经验特征与随机森林选取的特征保持有较好的一致性。另外，DEM、最大差异均值、绿色波段、红色波段、地表曲率、山体阴影、坡面曲率、相对高差、水流方向和坡向等特征对于提取农用地也具有一定指导作用和启发意义。

(8) 陡崖

陡崖特征选择结果如图 4-9 所示，实验共选出了 49 个特征，其中坡度贡献度最大，达到 19.35；贡献度最小的特征是绿色波段与邻接对象的差异，仅为 2.65。

所选特征集贡献度均值为 6.44，大于均值的特征包括：1.坡度均值；2.坡度均值灰度共生矩阵；3.植被指数均值；4.DEM 同质性灰度共生矩阵；5.最大差异均值；6.水流方向均值；7.山体阴影均值灰度共生矩阵；8.坡度与邻接对象的差异；9.山体阴影均值；10.近红外波段同质性灰度共生矩阵；11.坡向均值；12.红色波段均值灰度共生矩阵；13.绿色波段均值；14.坡向均值灰度共生矩；15.绿色波段均值灰度共生矩阵；16.绿色波段均值；17.坡向与影像对象主轴方向的夹角；18.地表曲率标准差；19.相对高差。

在滑坡信息提取中，根据专家知识设定的陡崖提取规则为：坡度 ≥ 45 ；植被指数 ≥ 0.12 ；亮度 ≤ 90 。其中，坡度和植被指数的贡献度排名靠前，分别为第一和第三，亮度排名靠后，仅为第二十二位。因此在陡崖提取中，专家经验特征与随机森林选取的特征保持有一定的一致性。另外，DEM、最大差异均值、水流方向、山体阴影、绿色波段、坡向、红色波段、近红外波段、坡向与影像对象主轴方向的夹角、地表曲率和相对高差等特征对于提取农用地也具有一定指导作用和启发意义。

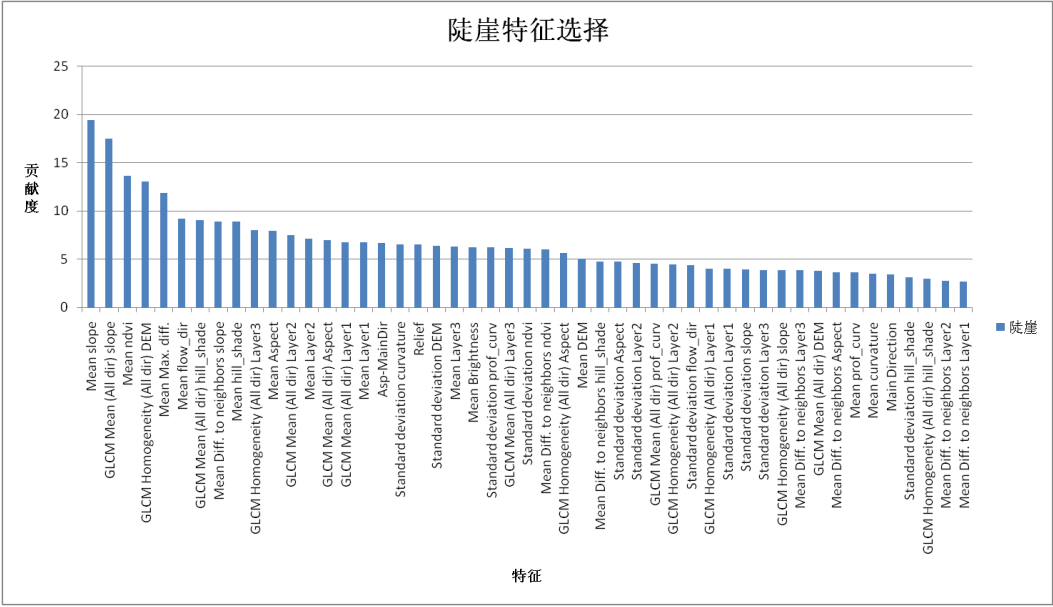


图 4-9 陡崖特征选择结果

(9) 道路

道路特征选择结果如图 4-10 所示，实验共选出了 3 个特征，分别为：1.宽度，贡献度 2.97；2.植被指数均值，贡献度 2.67；3.长度，贡献度 2.62。选取的特征贡献度均不高，特征集贡献度均值仅为 2.75。

在滑坡信息提取中，根据专家知识设定的道路提取规则为：长宽比 ≥ 3 ；坡向与主方向的夹角： $[80,100]$ 。由于该区域影像中道路对象较少，选取的样本不够典型，所以特征选择的结果中并没有“长宽比”和“坡向与主方向的夹角”特征。但是，长宽比特征是由对象的长度和宽度共同决定的，而这两个特征被选取出来了。因此在道路提取中，专家经验特

征与随机森林选取的特征仍保持有一定的一致性。另外，植被指数特征对于提取道路也具有指导作用和启发意义。

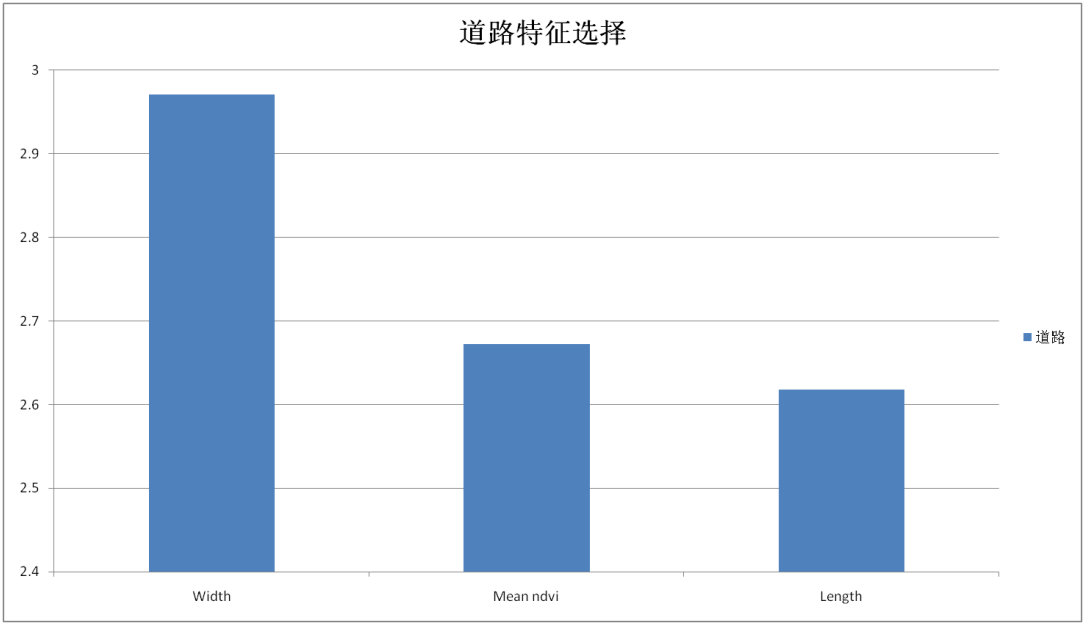


图 4-10 道路特征选择结果

（10）滑坡

本文采用的面向对象滑坡信息提取是通过逐步剔除滑坡易混淆地物后得到滑坡的过程，实验中并没有明确的设定专家经验特征。在这里通过对滑坡地物样本进行随机森林特征选择，得到的特征将用于指导直接提取滑坡的诊断性特征。

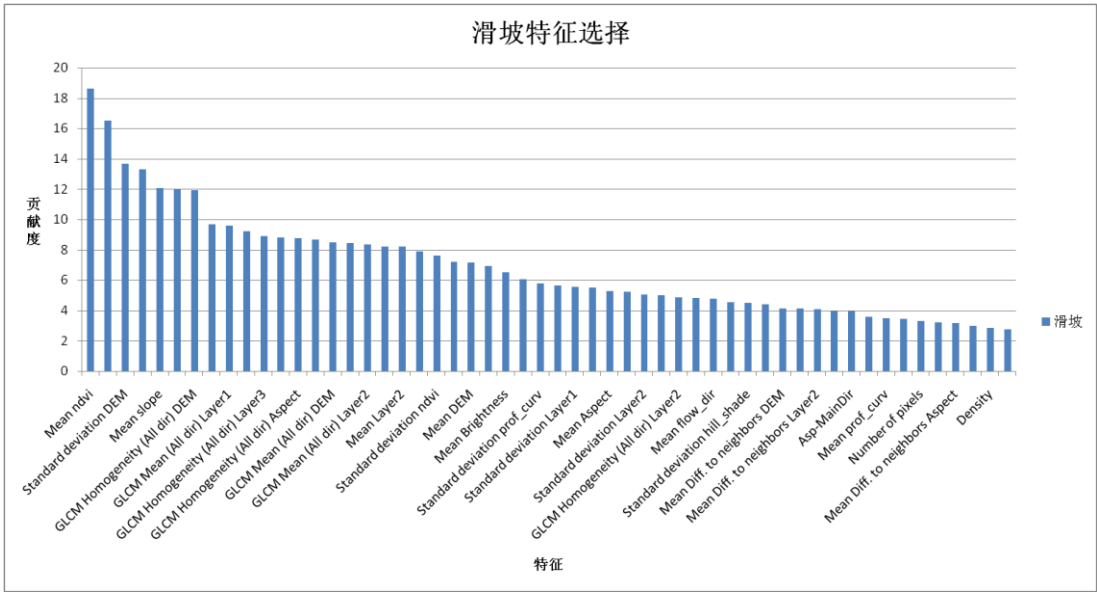


图 4-11 滑坡特征选择结果

滑坡地物特征选择结果如图 4-11 所示，实验共选出了 54 个特征，其中植被指数贡献

度最大，达到 18.62；贡献度最小的特征是坡面曲率均值灰度共生矩阵，仅为 2.76。

所选特征集贡献度均值为 6.90，大于均值的特征包括：1.植被指数均值；2.最大差异均值；3.DEM 标准差；4.相对高差；5.坡度均值；6.坡度均值灰度共生矩阵；7.DEM 同质性灰度共生矩阵；8.绿色波段均值；9.绿色波段均值灰度共生矩阵；10.坡向标准差；11.近红外波段同质性灰度共生矩阵；12.山体阴影均值灰度共生矩阵；13.坡向同质性灰度共生矩阵；14.山体阴影均值；15.DEM 均值灰度共生矩阵；16.近红外波段均值灰度共生矩阵；17.绿色波段均值灰度共生矩阵；18.近红外波段均值；19.绿色波段均值；20.水流方向标准差；21.植被指数标准差；22.地表曲率标准差；23.DEM 均值；24.坡度与邻接对象的差异。由上述特征选择结果可以发现，植被指数、最大差异均值、DEM、相对高差、坡度、绿色波段、坡向、近红外波段、山体阴影、水流方向和地表曲率等特征对于提取滑坡将具有一定指导作用和启发意义。

(11) 全局

将实验区中所有样本对象代入随机森林进行特征选择，所选取出的特征可以指导传统的影像分类方法设定标准近邻特征空间。

全局特征选择结果如图 4-12 所示，实验共选出了 63 个特征，其中坡度贡献度最大，达到 30.79；贡献度最小的特征是对象长轴长度特征，仅为 2.96。

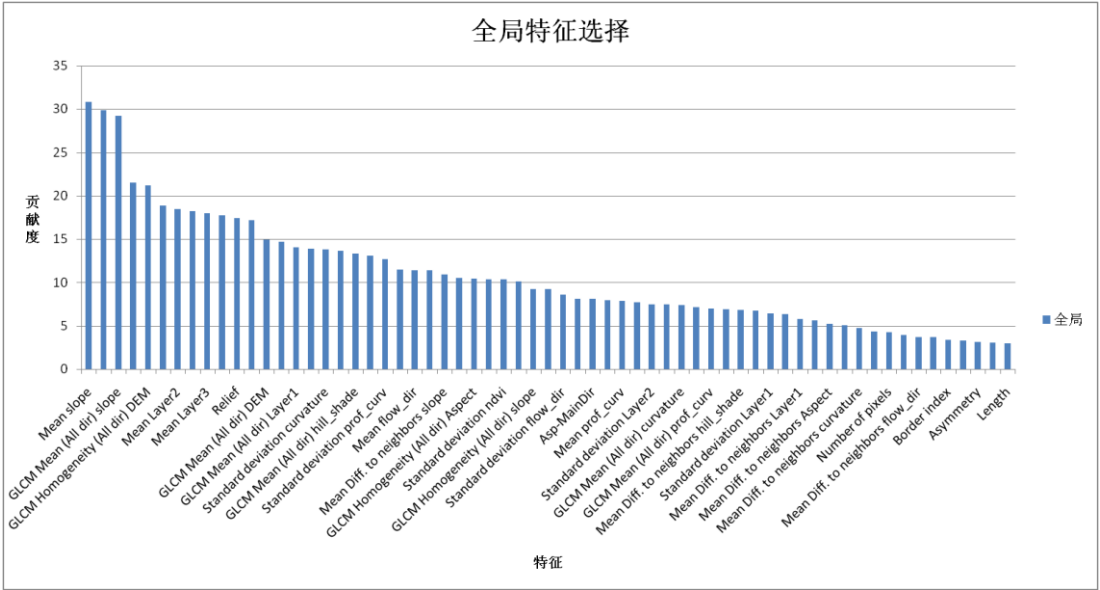


图 4-12 全局特征选择结果

所选特征集贡献度均值为 10.77，大于均值的特征包括：1.坡度均值；2.植被指数均值；3.坡度均值灰度共生矩阵；4.最大差异均值；5.DEM 同质性灰度共生矩阵；6.红色波段均值灰度共生矩阵；7.红色波段均值；8.近红外波段均值灰度共生矩阵；9.近红外波段均值；10.DEM 标准差；11.相对高差；12.DEM 均值；13.DEM 均值灰度共生矩阵；14.亮度均值；15.绿色波段均值灰度共生矩阵；16.绿色波段均值；17.地表曲率标准差；18.山体阴影均值；19.山体阴影均值灰度共生矩阵；20.近红外波段同质性灰度共生矩阵；21.坡面曲率标准差；22.植被指数与邻接对象的差异；23.水流方向；24.坡向标准差；25.坡度与邻接对象的差异。

由上述特征选择结果可以发现，坡度、植被指数、最大差异均值、DEM、红色波段、近红外波段、相对高差、亮度、绿色波段、地表曲率、山体阴影、坡面曲率、水流方向和坡向等特征对于在该区域进行传统影像分类设定标准近邻特征空间具有一定指导作用和启发意义。

§4.4 本章小结

本章主要研究随机森林滑坡特征选择方法。首先，介绍随机森林的基本原理，然后，提出基于对象的随机森林滑坡特征选择方法，最后，进行随机森林滑坡特征选择实验。随机森林滑坡特征选择方法，利用随机森林对初始滑坡信息提取特征进行选择，并计算得到所选取特征的贡献度。通过实验发现，经验选取的特征与随机森林选取的特征有着很强的一致性，且在当中有着很高的贡献度。该方法对于寻找滑坡信息提取潜在特征具有指导性作用。

第五章 面向对象滑坡信息提取中 DEM 空间分辨率影响分析

DEM 在滑坡信息提取中起到了重要作用,可在实际应用中往往难以获取到研究区的高分辨率 DEM 数据,这种情况下,分析 DEM 分辨率对滑坡提取的影响具有重要意义。本章首先建立由 DEM 及其衍生数据辅助高分遥感影像的面向对象滑坡信息提取流程,然后通过对 DEM 数据进行重采样,将其按空间分辨率的不同分为 7 组进行实验,最后分析 DEM 空间分辨率变化对滑坡信息提取的影响。研究发现,在无法获取实验区高分辨率 DEM 的情况下,使用分辨率较低的 DEM 产品也可实现滑坡探测的目的。

§5.1 研究区及数据描述

研究区位于印度北部的北阿肯德邦(Uttarakhand)地区乌基默特(Okhimath)市周围玛亚马赫什瓦河(Madhyamaheshwar)流域 28km² 的区域。该地区出露的岩石主要为花岗岩片麻岩、石英绢云母片岩、石英岩、石榴子石的云母片岩、大理岩和间歇的基性侵入岩,新构造运动导致岩石切割显著,有很多不同类型的滑坡发生,为滑坡提取提供了一个很好的训练场。1998 年 8 月,由于降雨引发的山体滑坡,仅在 Mandakini 山谷就有 466 处,造成了 103 人死亡,47 个村庄遭到破坏^[24]。

滑坡信息提取使用的实验数据来自荷兰屯特大学(University of Twente)地理信息科学与地球观测学院(Faculty of Geo-Information Science and Earth Observation, ITC)Martha 等在网上共享的示例数据(<http://www.itc.nl/OOA-group>),包括多光谱数据和 DEM 数据:1)多光谱数据是由搭载在 IRS-P6(Resourcesat-1)卫星上线性成像自扫描(LISS-IV)传感器于 2004 年 4 月 16 日获取^[68],包括近红外(0.76-0.86 μ m)、红(0.62-0.68 μ m)、绿(0.52-0.59 μ m)三个波段,星下点几何分辨率为 5.8m,主要用来参与影像的分割运算,以及提取光谱诊断性特征;2)DEM 数据由 IRS-P5(Cartosat-1)卫星于 2006 年 4 月 6 日获取的立体像对通过数字摄影测量的方法制作而成^[69],空间分辨率为 10m,垂直方向上均方根误差(RMSE)为 2.31m,主要用于产生坡度、坡向等 DEM 衍生数据。

§5.2 面向对象的滑坡信息提取方法

滑坡的信息提取方法主要源自于专家在影像解译时的滑坡探测知识,因此它模拟了滑坡专家在影像解译时的认知过程。本文在 Martha 提出的面向对象滑坡信息提取方法^[24]基础上加以改进,得到的面向对象滑坡信息自动提取的步骤如图 5-1 所示。滑坡识别和分类的

过程主要通过下面 4 步来描述：

- 1) 创建影像对象。采用多尺度分割方法，根据像元值的同质性将影像分成不同的区域。本文采用的分割尺度设为 10，形状参数为 0.1，紧致度为 0.9；
- 2) 确定候选滑坡对象。滑坡发生后裸露的岩石和碎屑被暴露出来，这使得它与山区的植被有很好的区分度，因此我们用 NDVI 作为第一个准则来确定候选对象；
- 3) 剔除易混地物。通过 NDVI 特征提取出候选对象中混有阴影、水体、滩涂和裸露岩石等易混地物，利用光谱、形态和上下文信息通过面向对象分析方法逐步剔除它们即可得到滑坡对象；
- 4) 确定滑坡类型。美国学者 Varnes 根据不同滑坡的物质成分和移动形式，以滑坡规模、滑面形状、干湿条件进行三维分布式分类^[70]。实验针对研究区的特点，选取了浅层平移型岩石滑坡、碎石滑坡、碎石流、旋转型岩石滑坡、平移型岩石滑坡 5 类典型的滑坡进行滑坡对象分类。



图 5-1 面向对象滑坡信息提取

面向对象的滑坡信息提取过程在 eCognition 软件中完成。除了文章第 5.2 节提到的多光谱数据和 DEM 数据外，滑坡信息提取过程中还使用到了由 ArcGIS 软件计算到的 DEM 衍生数据，包括：坡度、坡向、山体阴影、地表曲率、剖面曲率、河网等级等。将原始 5.8 米分辨率多光谱遥感影像和 10 米分辨率 DEM 数据及其衍生数据一同导入 eCognition 软件，采用表 5-1 中的特征规则对该研究区进行面向对象的滑坡信息提取，得到图 5-2(b)提取结果。

重采样 DEM 可得到不同空间分辨率的 DEM 数据,是获取不同空间分辨率数据的重要方法^[71]。为研究 DEM 对面向对象滑坡信息提取精度的影响,需要对原始的 DEM 数据进行重采样。重采样会削弱原始 DEM 数据的特征,同时也会影响 DEM 衍生数据(坡度、地表曲率等)的准确性^{[71][72]},最近邻(NEAR)方法影响较大,双线性内插(BIL)和三次卷积(CUB)影响较小。本实验使用对数据影响较小的双线性内插方法将 10m 分辨率的 DEM 重采样为 15m、20m、30m、50m、75m、100m 来研究 DEM 空间分辨率对滑坡信息提取的影响。

首先保持多光谱影像不变,将重采样后的不同分辨率 DEM 及其衍生数据替代原数据,然后采用表 5-1 中所列的提取规则进行滑坡信息提取。由于重采样后的 DEM 特征发生变化,得到的提取效果并不理想,存在许多误分类现象,部分滑坡对象因为 DEM 特征的变化而未能识别出来,需要针对不同分辨率 DEM 对每种地物的提取参数值做相应的调整。调整后的参数值列在表 5-2 中,得到的最终分类结果如图 5-2 所示。

表 5-1 滑坡信息提取分类规则

地物类型	特征规则	评判依据
阴影	1. 山体阴影≤92; 2. 植被指数≥0.1。	山体阴影是利用多光谱影像获取时的太阳方位,由 DEM 产生的表示阴影状态的虚拟影像;植被指数条件使得被阴影覆盖的滑坡对象不随阴影一同被剔除。
主流	1. 河网等级≥5。	利用 DEM 自动生成水系图并用 Strahler 方法 ^[73] 对它们分级,本研究区域主流河网的河网等级≥5。
支流	1. 长宽比≥3.6; 2. 坡度≤17.5; 3. 植被指数≤0.067; 4. 近红外波段≤74。	河流对象表现为细长状,长宽比高;河流相对平缓,坡度值不高;河流中几乎不含植被,植被指数低;水体对近红外波段有吸收作用,近红外波段值低。
湖泊	1. 近红外波段≤55; 2. 坡度≤5。	湖泊中水体对近红外波段有吸收作用,近红外波段值低;湖面平坦,坡度值低。
主流滩涂	1. 亮度≥65; 2. 与主流的距离≤100m; 3. 坡度≤20; 4. 相对高差≤30。	主流滩涂对可见光反射率高,表现为高亮度值;通常与河流相邻接,与主河流距离近;相对平缓,坡度值低,相对高差小。
支流滩涂	1. 植被指数≤0.08; 2. 坡度≤15; 3. 相对高差≤20。	支流滩涂植被较少,植被指数低;相对平缓,坡度值低,相对高差小。
建筑区	1. 坡度≤12; 2. 紧致度<2.5; 3. 红色波段同质性灰度共生矩阵: [0.15,0.2]。	建筑通常修在平缓地区,坡度值低;建筑图案形成的典型纹理使得它与邻接对象有很高的标准差,根据紧致度和灰度共生矩阵可将建筑区提取出来。
农用地	1. 坡度≤30; 2. 植被指数≥0.094; 3. 红色波段均值灰度共生矩阵: [60,90]。	中低坡度的研究区域大多已经转变成适合耕作的梯田;梯田中的植被虽没有林地那样密集,但仍有较高的植被指数;研究区的梯田与等高线平行,且宽度大都一致,在影像上表现为独特的纹理特征,通过红色波段均值灰度共生矩阵均值能清楚的区分梯田图案。
树木	1. 坡度≤30; 2. 植被指数≥0.094; 3. 红色波段均值灰度共生矩阵: [55,60]。	部分在之前过程中为分离出的树木,纹理特征与农用地不同,通过红色波段均值灰度共生矩阵均值可区分其图案。

地物类型	特征规则	评判依据
岩区	1. 亮度 ≤ 95 ; 2. 植被指数 ≥ 0.123 ; 3. 坡度: $[29.5, 45]$ 。	岩区地表粗糙度没有新生的滑坡高, 所以它仅具有中等亮度值; 岩区偶尔也会混有少许植被, 植被指数偏高一点; 岩石区域主要分布在研究区的中等坡度范围。
陡崖	1. 亮度 ≤ 90 ; 2. 植被指数 ≥ 0.12 ; 3. 坡度 ≥ 45 。	陡崖相对于岩区更陡峭, 主要分布在研究区的高等坡度范围。
道路	1. 坡向与主方向的夹角: $[65, 115]$; 2. 长宽比 ≥ 4.5 ; 3. 近红外波段 ≥ 76.5 。	坡向是坡度下降最快的方向, 道路(山区公路)通常沿着与坡向垂直的方向修建, 所以道路对象的主方向与坡向几乎垂直; 道路表现为细长状, 长宽比高; 道路对近红外波段反射强, 近红外波段反射值高。
浅层平移型岩石滑坡	1. 不对称性 ≥ 0.95 ; 2. 坡度 ≥ 25 。	浅层平移型岩石滑坡发生在地球表层, 滑坡外围到中轴线的距离很短, 表现为较强的不对称性; 发生在较陡峭的区域, 坡度值较高。
碎石滑坡	1. 与农用地的邻接比 ≥ 0.5 。	碎石滑坡一般发生在严重风化区或地形表面覆盖深度很大的区域, 通常与农用地相邻接。
碎石流	1. 坡度 ≥ 25 ; 2. 长度 ≥ 500 。	含水的碎石在空间上持续运动形成碎石流, 通常具有中等程度的坡度和很长的运行区, 并且冲刷作用能沿着运行路径明显地观察到。
旋转型岩石滑坡	1. 地表曲率 ≤ -1 。	运动的岩石沿着斜坡滑下后, 旋转型岩石滑坡的破裂处表现为弯曲或上凹的表面, 表现为坡度形态的突然变化, 地表曲率 ≤ -1 。
平移型岩石滑坡	1. 坡面曲率 ≤ 1 。	运动的岩石沿着斜坡滑下后, 平移型岩石滑坡的破裂处表现为平坦或波浪状表面。地表曲率值非常低, 坡面曲率 ≤ 1 , 有时接近为零。

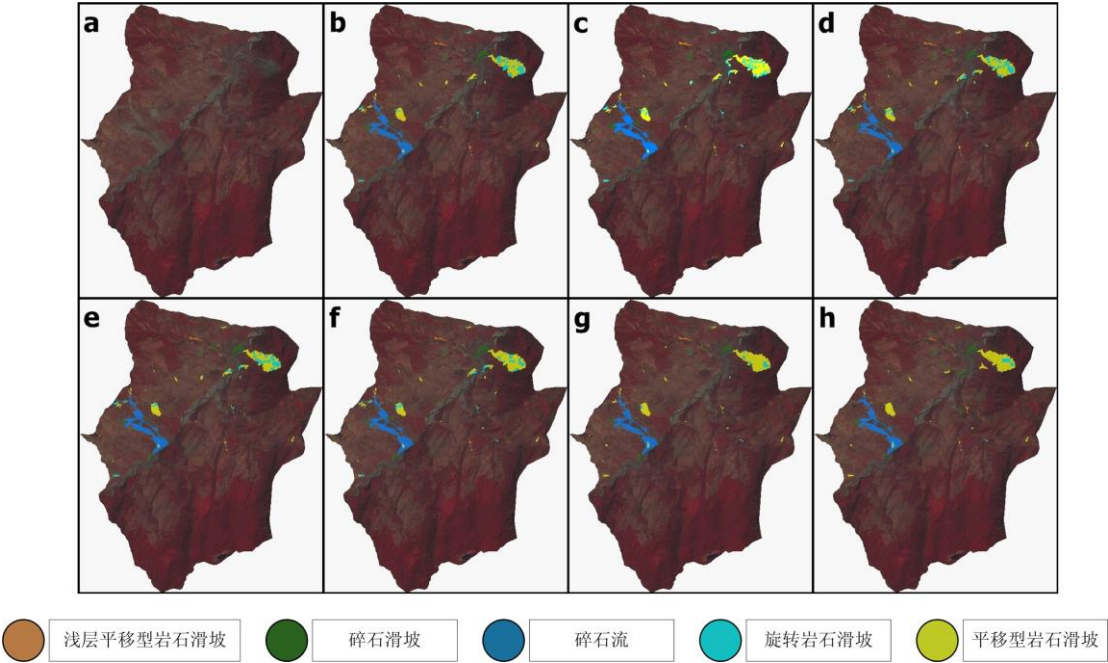


表 5-2 滑坡信息提取参数设定

地物类型	特征规则	特征参数值						
		10m	15m	20m	30m	50m	75m	100m
支流	坡度	≤17.5	≤17.5	≤18.5	≤18.5	≤18.5	≤18.5	≤18.5
湖泊	坡度	[0, 5]	[0, 5]	[0, 5]	[0, 5]	[0, 5.9]	[0, 5.9]	[0, 5.9]
主流滩涂	坡度	[0, 20]	[0, 20]	[0, 19.75]	[0, 16]	[0, 16]	[0, 16]	[0, 16]
	相对高差	≤30	≤38.95	≤38.95	≤38.95	≤38.95	≤38.95	≤38.95
支流滩涂	相对高差	≤20	≤18.4	≤18.4	≤18.4	≤18.4	≤18.4	≤18.4
农用地	坡度	[0, 30]	[0, 29.5]	[0, 29.5]	[0, 29.5]	[0, 29.5]	[0, 29.5]	[0, 29.5]
树木	坡度	[0, 30]	[0, 29.5]	[0, 29.5]	[0, 29.5]	[0, 29.5]	[0, 29.5]	[0, 29.5]
岩区	坡度	[29.5,	[29.5,	[29.5,	[29.5,	[29.5,	[29.5,	[29.5,
		45]	44.5]	44.5]	44.5]	44.5]	44.5]	44.5]
陡崖	坡度	≥45	≥44.5	≥44.5	≥44.5	≥44.5	≥44.5	≥44.5
道路	坡向与主方向	[65,	[65,115]	[65,115]	[65,115]	[65,115]	[65,115]	[63.5,115]
	夹角	115]						
浅层平移型岩石	坡度	≥25	≥25	≥25	≥25	≥25	≥25	≥22.5
滑坡								
碎石滑坡	与农用地的邻	≥0.5	≥0.4	≥0.4	≥0.4	≥0.4	≥0.4	≥0.4
	接比							
碎石流	坡度	≥25	≥25	≥25	≥25	≥25	≥24	≥24

§5.3 实验结果及分析

5.3.1 滑坡信息提取结果

从图 5-2 中的结果中可以看出，在提取参数值做了调整的情况下，研究区内的滑坡基本能被探测出来。随着 DEM 分辨率降低会出现许多滑坡类型的误分现象，主要表现在研究区东北方一个较大面积滑坡（图 5-3 中方框内的区域）的分类中。下面针对这一区域滑坡的分类情况做详细的精度分析，从中发现 DEM 空间分辨率变化对滑坡信息提取的影响。

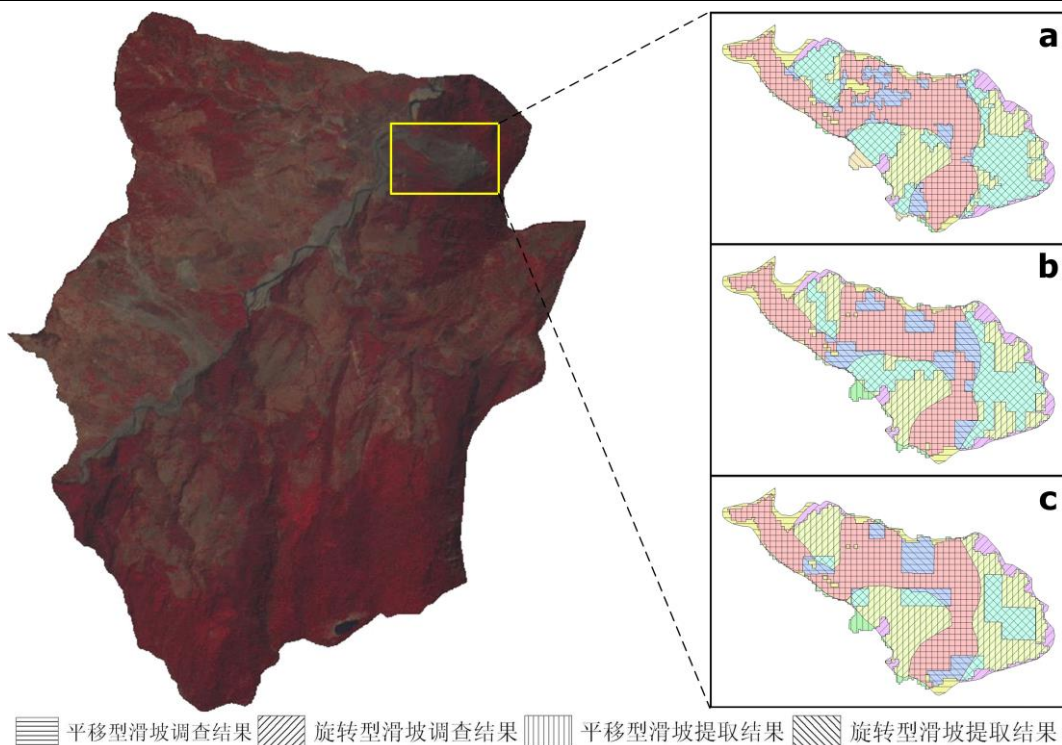


图 5-3 不同空间分辨率 DEM 提取滑坡对比图

(a)10m 分辨率, (b)30m 分辨率, (c)50m 分辨率

5.3.2 精度评价

选取 10m、30m 和 50m 分辨率的分类结果与 1998 年 8 月发生大型滑坡后制作的滑坡调查结果^[24]相叠加，得到评价结果对比图（如图 5-3 所示）。图中调查结果与提取结果的填充符号设为相互正交的直线，因此直线相交的区域为正确识别的滑坡（即提取结果和调查结果均为滑坡），直线未相交的区域表示未识别和多识别的滑坡。在直线相交区域中直线正交表示正确分类的滑坡（即提取结果和调查结果的滑坡类型相同），直线斜交表示错误分类的滑坡（即提取滑坡类型与调查滑坡类型不一致）。通过将不同分辨率的分类结果与人工调查结果进行比较，对正确识别的滑坡、未识别的滑坡和多识别的滑坡进行统计分析，得到图 5-4(a)所示的精度评价结果。对正确识别滑坡中正确分类的滑坡和错误分类的滑坡进行统计分析，得到图 5-4(b)所示的精度评价结果。各个统计值的计算公式如下：

$$\text{正确识别的滑坡} = \frac{\text{正确识别的滑坡总面积}}{\text{调查的滑坡总面积}} \quad (5-1)$$

$$\text{未识别的滑坡} = \frac{\text{未识别的滑坡总面积}}{\text{调查的滑坡总面积}} \quad (5-2)$$

$$\text{多识别的滑坡} = \frac{\text{多识别的滑坡总面积}}{\text{调查的滑坡总面积}} \quad (5-3)$$

$$\text{正确分类的滑坡} = \frac{\text{正确分类的滑坡总面积}}{\text{正确识别的滑坡总面积}} \quad (5-4)$$

$$\text{错误分类的滑坡} = \frac{\text{错误分类的滑坡总面积}}{\text{正确识别的滑坡总面积}} \quad (5-5)$$

5.3.3 结果分析及结论

随着 DEM 空间分辨率的降低，滑坡信息提取结果会逐渐受到影响。通过对提取过程中每种地物的特征参数值做相应的调整，可以最大程度的减少误分类现象。从表 5-2 中参数调整情况我们可以看出，变化的特征规则主要包括：坡度、相对高差、坡向与主方向夹角、与农用地的邻接比例，其中“与农用地的邻接比例”表示影像对象与邻接的农用地对象重合边界的长度，而农用地类型的确定需要用到坡度特征。从这些特征规则中我们可以发现，它们之间的共同特点是都与 DEM 相关，DEM 分辨率的变化导致了这些特征值相应的变化。

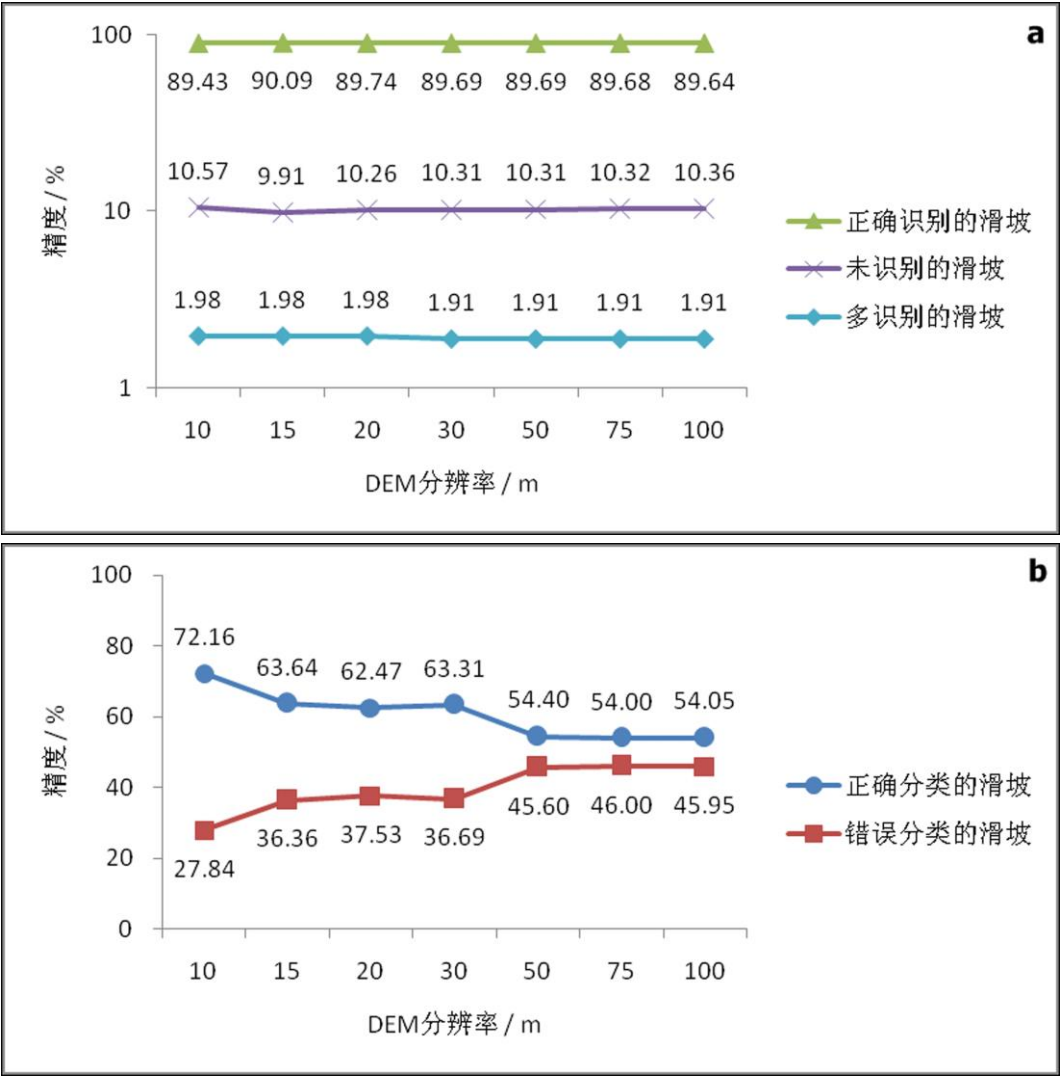


图 5-4 滑坡信息提取精度
(a)滑坡识别精度, (b)滑坡分类精度

通过比较图 5-2 中不同空间分辨率 DEM 参与的滑坡信息提取结果可以看出, 图 5-2(h) 低分辨率 DEM 提取结果与图 5-2(b) 高分辨率 DEM 提取结果相比除了少数小面积的滑坡不能探测出来外, 大面积的滑坡都可以被识别。在图 5-4(a) 中反映为, 随着 DEM 分辨率的变化, 正确识别的滑坡精度始终保持在 90% 左右。这意味着, 在 DEM 辅助的滑坡信息提取中, 通过设置恰当的提取规则, 可以实现滑坡探测的目的。

在滑坡类型的识别方面, 从图 5-4(b) 中可以看出, 当 DEM 分辨率为 10m 时, 滑坡的分类精度达到 72.16%。随着 DEM 分辨率的降低, 滑坡的分类精度也逐渐降低, 与此同时滑坡类型的误分现象也逐渐增多。当 DEM 分辨率降低到 50m 时, 滑坡分类精度降低到 60% 以下。当 DEM 分辨率继续降低滑坡分类精度保持在 54% 左右。

由此可见, DEM 分辨率的精细程度对面向对象滑坡信息提取有较大影响。对于滑坡识别方面的影响 (图 5-4a), 通过调整 DEM 特征相关的参数, 仍可保持较好的识别精度。对于滑坡类型的分类精度 (图 5-4b), 会随 DEM 分辨率下降而降低。造成这一现象的主要原因是区分旋转型和平移型岩石滑坡的过程中需要通过地表曲率 (curvature) 特征进行分割, 而地表曲率是由 DEM 计算得到, 与 DEM 分辨率相同。研究区内滑坡面积在 5000 至 40000m² 之间, 对于面积大于 5000m² 的滑坡, 当 DEM 分辨率为 10m 时能够得到较好的分类结果。当 DEM 分辨率在 10~30m 之间时, 可以勉强区分出滑坡类型。当 DEM 分辨率大于 30m 时, 旋转型滑坡仅表示为几个像素, 不能得到很好的分类结果。

通过上述实验, DEM 空间分辨率对面向对象滑坡信息提取的影响可以概括为以下 4 个方面:

- 1) 高空间分辨率 DEM 辅助的面向对象滑坡信息提取能得到较高的滑坡分类精度;
- 2) 随着 DEM 空间分辨率的降低, 滑坡信息提取结果会逐渐受到影响, 但通过调整 DEM 特征相关的参数, 可以实现滑坡探测的目的;
- 3) 当 DEM 空间分辨率小于 30m 时, 对面积大于 5000m² 的滑坡能得到较好的识别和分类结果;
- 4) 当 DEM 空间分辨率大于 30m 时, 旋转型滑坡仅表示为几个像素, 不能得到很好的分类结果。

§5.4 本章小结

本章通过对比实验, 分析了 DEM 空间分辨率的变化对面向对象滑坡信息提取的影响。首先采用 DEM 及其衍生数据辅助高分遥感影像, 实现综合光谱、形状和地形特征的面向对象滑坡信息提取。然后将 DEM 重采样为 7 种分辨率等级分别进行滑坡信息提取, 针对不同分辨率 DEM, 对每种地物的提取参数值做相应的调整, 得到各自分辨率下的滑坡提取结果。最后将不同分辨率 DEM 下的提取结果与调查结果相叠加, 评价提取精度, 分析 DEM 空间分辨率变化对高分影像面向对象滑坡信息提取的影响。实验结果表明, 当 DEM 分辨率小于 30m 时, 对面积大于 5000m² 的滑坡能得到较好的识别和分类结果; 大于 30m 时虽难以区分出滑坡类型, 但通过调整部分规则参数仍能实现对滑坡的探测。本研究对于准确提取滑坡信息所需 DEM 产品的选择具有重要的指导意义和参考价值。

第六章 结论与展望

§6.1 结论

本文对面向对象遥感滑坡信息提取技术进行了研究,针对分割参数如何设置,分类特征如何确定,DEM 产品如何选择这三个问题进行了深入探讨,具体研究结论体现在以下几个方面:

1) 基于 SSC (Scale-Shape-Compactness) 的滑坡对象构建方法。提出一种 SSC 多尺度分割参数优化方法,通过对湖南省资兴市滑坡灾害区进行 SSC 分割参数优化实验,得到尺度 (Scale)、形状因子 (Shape) 和紧致度 (Compactness) 优化分割参数 Seg_{opt} , 将优化分割参数 Seg_{opt} 与默认分割参数 Seg_{def} 进行对比验证。通过实验结果得出,基于 SSC 的滑坡对象构建方法,在不具备先验知识的前提下,可以计算出一个分割参数,使得分割后产生的滑坡对象与实际目标轮廓最为接近。该方法对于面向对象遥感滑坡信息提取中如何设置多尺度分割参数具有重要的指导意义和应用价值。

2) 随机森林滑坡特征选择方法。提出一种基于对象的随机森林滑坡特征选择方法,通过 R 语言编写程序,对印度北部的北阿肯德邦地区滑坡信息提取过程中的地物进行随机森林特征选择,并计算得到所选取特征的贡献度。通过实验分析发现,利用随机森林特征选择方法得到的特征与根据专家经验选出的特征有着很强的一致性,且专家经验特征在随机森林选出的特征集中有着很高的贡献度。该方法在面向对象遥感滑坡信息提取中,对于寻找地物潜在诊断性特征具有重要的参考价值和启发意义。

3) 面向对象滑坡信息提取中 DEM 空间分辨率影响分析。为分析 DEM 分辨率对滑坡提取的影响并确定满足应用的分辨率要求,将原始 10m 分辨率 DEM 重采样为空间分辨率不同的 15m、20m、30m、50m、75m、100m,采用 DEM 及其衍生数据辅助高分遥感影像的面向对象滑坡信息提取方法进行实验。通过实验分析得出,当 DEM 分辨率小于 30m 时,对面积大于 5000m² 的滑坡能得到较好的识别和分类结果;大于 30m 时虽难以区分出滑坡类型,但通过调整部分规则参数仍实现对滑坡的探测。该研究对于准确提取滑坡信息所需 DEM 产品的选择具有重要的指导意义和参考价值。

本文的创新之处有:

1) 提出了一种基于 SSC (Scale-Shape-Compactness) 的滑坡对象构建方法。该方法针对面向对象滑坡信息提取多尺度分割中分割参数设置问题,通过结合影像对象内均质度、对象间异质度以及对象面积创建的分割质量评价指标,分别从多尺度分割的尺度参数、形状因子参数和紧致度参数三方面进行优化,以得到面向对象遥感滑坡信息的最优多尺度分割参数。

2) 提出了一种基于随机森林的面向对象滑坡特征选择方法。该方法针对面向对象遥感滑坡信息提取中特征选择问题,结合随机森林在特征选择方面的优势,通过 R 语言编写

的程序，分别对遥感滑坡信息提取中各类地物进行随机森林特征选择运算，得到各个特征的贡献度，以指导目标地物的诊断性特征的选取。

§6.2 展望

本文对面向对象影像分析技术在滑坡信息提取中的应用进行了详细的研究，通过作者的亲身实践，注意到有些内容存在着一些有待提高或解决的问题，主要表现在：

（1）探索新的分割参数优化方法

本文提出的 SSC 多尺度分割参数优化方法存在一定的局限性，当优化尺度参数时，形状因子和紧致度采用的是默认值，在一定程度上会影响尺度参数选择，无法做到真正意义上的全局最优。考虑如何实现真正意义上全局最优的分割参数优化方法。

（2）设计自动设定特征阈值的方法

在面向对象滑坡信息提取中，当地物的诊断性特征确定后，还需要设定特征阈值才能将目标地物准确地提取出来，而特征阈值设定大多通过反复尝试或根据专家经验得到。可尝试研究一种方法，通过选取有限的地物样本，自动分析出特征的阈值范围，为滑坡信息提取提供参考，以减少人工尝试的步骤，提升滑坡信息提取的工作效率。

（3）寻找更多有效的地物信息提取诊断性特征

由于考虑到运算效率和适用性的关系，本文的特征选择实验中仅挑选了 76 个候选特征，可能存在其他对地物信息提取贡献度大的特征，由于未列入候选特征而没有被选择出来。尝试引入更多候选特征在更多不同信息提取环境中进行特征选择实验，以探索更多有限的地物信息提取诊断性特征。

（4）分析地物特征的物理意义

运用特征选择方法得到的贡献度大的地物特征，通常对应于某种物理意义。如，滑坡地物 NDVI 贡献度较高，是因为滑坡的发生破坏了山上的植被；湖泊地物的相对高差贡献度较高，是由于湖泊表面平坦，高程值接近。通过分析地物特征的物理意义，可以更加深入的了解地物的特性，启发面向对象信息提取过程。

致 谢

时光飞逝，转眼间在中国地质大学（武汉）的学习生活即将结束，回想一路走过的求学之路，我感慨万千。在地大的学习生涯令我毕生难忘，在这里师长、同学、亲人和朋友都给过我无私的帮助与关怀，在我的内心里永远存有感激之情。

值此论文完成之际，首先要感谢我的导师刘修国老师。刘老师广博深厚的理论素养、活跃创新的学术思维、严谨扎实的治学态度和对学术事业一丝不苟、孜孜不倦的追求精神都给我留下了极深的印象，使我受益匪浅。从论文选题、提纲组织、写作方法到后期的审阅修改，无不倾注了刘老师的智慧与汗水。在此，我要对我的恩师表示衷心的感谢和崇高的敬意！

感谢陈启浩老师对我工作、学习以及生活上的关心和支持！您教会我许多专业知识与编程技术，也给过我许多帮助和关心。您严谨的逻辑思维、一丝不苟、积极向上的工作作风和科研态度，永远是我学习的榜样，您对我生活学习上的关心和支持我永远铭记在心。

感谢信工学院遥感系的全体老师们，谢谢你们将我带入了遥感的学术世界。忘不了在中地数码公司实习时高伟、张唯、王红平、花卫华老师对我实践技能的指导与帮助；忘不了刘福江老师对我初次尝试科技论文写作时的启蒙与指导；忘不了宋妍、田玉刚、徐世武、肖启芝老师给我们带来的既充实又快乐的遥感专业课程；也忘不了在撰写项目实施方案时许凯、孙杰老师对我的指导与帮助。你们的支持和帮助伴随了我整个大学阶段的学习生活，在此我要衷心的说一声：老师，您辛苦了！

感谢沈永林、刘恒星、黄友昕、行晓黎、何浩、陈敬周、黄晓东、王孟琪、熊倩、邵芳芳、张敏、陈奇、姜萍、王晓威、杨帅、刘世永、徐乔、胡子林、邓瑾、任泽、黄玉青、杨准、鲁田、姚瑶、肖旻珣等实验室的师兄师姐学弟学妹们，谢谢你们在学习生活、科研实践和论文的开展过程中对我的帮助和启发。

感谢在大学里与我一同学习和生活的好朋友们：戢锐、孙飞、王珏、刘旭东、周长城、梁栋、马邵忠、柳凤仙、白洋、刘帅、殷强、刘良亮、黄博、张长兴、李登朝、吴伟龙等。感谢你们给予我生活和学习上的帮助，与你们融洽的相处使我紧张的学习生涯充满了乐趣，和你们一起奋斗的日子令我毕生难忘！

感谢我的父母及家人长期以来对我的鼓励、支持、理解和关心。没有你们的辛勤努力，就没有我的今天，我的每一步成功都离不开你们的默默支持与奉献。你们的期盼、关心与支持是我不断前进的动力。

最后把这篇文章献给所有曾经帮助、关心、支持我的人，谢谢大家！

参考文献

- [1] 王治华. 滑坡遥感调查、监测与评估[J]. 国土资源遥感, 2007, 71 (1):10-15, 23.
- [2] 李德仁, 眭海刚, 单杰. 论地理国情监测的技术支撑[J]. 武汉大学学报·信息科学版, 2012, 37(5):505-512.
- [3] 童立强, 郭兆成. 典型滑坡遥感影像特征研究[J]. 国土资源遥感, 2013, 25(1):86-92.
- [4] 黄慧萍, 吴炳方, 李苗苗等. 高分辨率影像城市绿地快速提取技术与应用[J]. 遥感学报, 2004, 8(1):68-74.
- [5] Singhroy V, Mattar K E, Gray A L. Landslide characterisation in Canada using interferometric SAR and combined SAR and TM images [J]. Advances in Space Research, 1998, 21:465-476.
- [6] Van Den, Eeckhaut M, et al. Use of LIDAR-derived images for mapping old landslides under forest [J]. Earth Surface Processes and Landforms, 2007, 32:754-769.
- [7] 王治华. 滑坡、泥石流遥感回顾与新技术展望[J]. 国土资源遥感, 1999, 41:10-15.
- [8] 宫鹏, 黎夏, 徐冰. 高分辨率影像解译理论与应用方法中的一些研究问题[J]. 遥感学报, 2006, 10:1-5.
- [9] Van Westen C J, Castellanos E, Kuriakose S L. Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview [J]. Engineering Geology, 2008, 102:112-131.
- [10] 王治华, 徐起德, 徐斌. 岩门村滑坡高分辨率遥感调查与机制分析[J]. 岩石力学与工程学报, 2009, 28:1810-1818.
- [11] 赵祥, 李长春, 苏娜. 滑坡泥石流的多源遥感提取方法[J]. 自然灾害学报, 2009, 18:29-32.
- [12] 邓辉, 巨能攀, 向喜琼. 高分辨率卫星遥感数据在白衣庵滑坡调查研究中的应用[J]. 地球与环境, 2005, 33:92-96.
- [13] 叶润青, 邓清禄, 王海庆. 基于图像分类方法滑坡识别与特征提取[J]. 工程地球物理学报, 2007, 4(6):574-577.
- [14] Borghuis A M, Chang K, Lee H Y. Comparison between automated and manual mapping of typhoon-triggered landslides from SPOT-5 imagery [J]. International Journal of Remote Sensing, 2007, 28:1843 - 1856.
- [15] Nichol J, Wong M S. Satellite remote sensing for detailed landslide inventories using change detection and image fusion [J]. International Journal of Remote Sensing, 2005, 26: 1913-1926.
- [16] Pike R J. The geometric signature: quantifying landslide - terrain types from digital elevation models [J]. Mathematical Geology, 1988, 20:491-511.
- [17] Iwahashi J, Pike R J. Automated classifications of topography from DEMs by an

- unsupervised nested-means algorithm and a three-part geometric signature [J]. *Geomorphology*, 2007, 86:409-440.
- [18] 姚鑫, 张永双, 王献礼等. 基于地貌特征的浅层崩滑体遥感自动识别[J]. *地质通报*, 2008, 27:1870-1874.
- [19] Barlow J, Martin Y, Franklin S E. Detecting translational landslide scars using segmentation of Landsat ETM+ and DEM data in the northern Cascade Mountains, British Columbia[J]. *Canadian Journal of Remote Sensing*, 2003, 29:510-517.
- [20] Baatz M, Schape A. Multiresolution segmentation: an optimization approach for high quality multi-scale image segmentation [C]. In: L.J. Strobl, T. Blaschke, T. Griesebener, *Angewandte geographische informations veraarbeitung XII, Beitrage zum AGIT Symposium Salzburg 2000*. Herbert Wichmann Verlag, Heidelberg, 2000, 12-23.
- [21] Moine M, Puissant A, Malet J P. Detection of landslides from aerial and satellite images with a semi-automatic method. Application to the Barcelonnette basin [C]. In: J. P. Malet, A. Remaitre, T. Bogaard, *Landslide Processes: From Geomorphological Mapping to Dynamic Modelling*. CERG, Strasbourg, France, 2009, 63-68.
- [22] Iwahashi J, Pike R J. Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature [J]. *Geomorphology*, 2007, 86(3):409-440.
- [23] 胡德勇, 李京, 赵文吉等. 基于对象的高分辨率遥感图像滑坡检测方法 [J]. *自然灾害学报*, 2008, 17(6):42-46.
- [24] Martha T R, Kerle N, Jetten V, et al. Characterising spectral, spatial and morphometric properties of landslides for semi-automatic detection using object-oriented methods [J]. *Geomorphology*, 2010, 116(1-2):24-36.
- [25] 沈永林, 李晓静, 吴立新. 基于航空影像和 LiDAR 数据的海地地震滑坡识别研究 [J]. *地理与地理信息科学*, 2011, 27(1):16-20.
- [26] 黄惠萍. 面向对象影像分析中的尺度问题研究[D]. 中国科学院研究生院博士学位论文, 2003:103-112.
- [27] 韩鹏, 龚健雅, 李志林等. 遥感影像分类中的空间尺度选择方法研究 [J]. *遥感学报*, 2010, 14(3):507-518.
- [28] Otsu N. A threshold selection method from gray-level histograms [J]. *IEEE transactions on systems, man, and cybernetics*, 1979, 9(1):62-66..
- [29] Benz U C, Hofmann P, Willhauck G, et al. Multi-resolution, object-oriented fuzzy analysis of remote sensing data for GIS-ready information [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2004, 58(3/4):239-258.
- [30] Definiens AG, eCognition Developer 8, Reference Book, Munich, Germany 2009.
- [31] Baatz M, Schäpe A. Multiresolution segmentation: An optimization approach for high quality multi-scale image segmentation. *Angewandte Geographische Informationsverarbeitung XII, Beitrage zum AGIT Symposium Salzburg, 2000*, 12-23.
- [32] Holt A C, Seto E Y W, Rivard T, et al. Object-based detection and classification of vehicles from high-resolution aerial photography [J]. *Photogrammetric Engineering and Remote*

- Sensing, 2009, 75(7):871-880.
- [33] Dragut L, Blaschke T. Automated classification of landform elements using object-based image analysis [J]. *Geomorphology*, 2006, 81(3/4):330-344.
- [34] Barlow J, Franklin S, Martin Y. High spatial resolution satellite imagery, DEM derivatives, and image segmentation for the detection of mass wasting processes [J]. *Photogrammetric Engineering and Remote Sensing*, 2006, 72(6):687-692.
- [35] Kerle N, de Leeuw J. Reviving legacy population maps with objectoriented image processing techniques [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2009, 47(7):2392-2402.
- [36] Laliberte A S, Rango A, Havstad K M, et al. Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico [J]. *Remote Sensing of Environment*, 2004, 93(1/2):198-210.
- [37] 柏延臣, 王劲峰. 遥感信息不确定性研究:分类与尺度效应模型[M]. 地质出版社, 2003.
- [38] Möller M, Lymburner L, Volk M. The comparison index: A tool for assessing the accuracy of image segmentation [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2007, 9(3):311-321.
- [39] Zhang H, Fritts J E, Goldman S A. Image segmentation evaluation: A survey of unsupervised methods [J]. *Computer Vision and Image Understanding*, 2008, 110(2):260-280.
- [40] Dragut L, Tiede D, Levick S R. ESP: A tool to estimate scale parameter for multiresolution image segmentation of remotely sensed data [J]. *International Journal of Geographical Information Science*, 2010, 24(6):859-871.
- [41] Espindola G M, Camara G, Reis I A, et al. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation [J]. *International Journal of Remote Sensing*, 2006, 27(14):3035-3040.
- [42] Blaschke T, Burnett C, Pekkarinen A. Image segmentation methods for object-based analysis and classification [J]. *Remote Sensing Image Analysis*, 2006:211-236.
- [43] Breiman L, Random forests. *Machine Learning*, 2001, 45:5-32.
- [44] Watts J D, Lawrence R L, Miller P R, et al. Monitoring of cropland practices for carbon sequestration purposes in north central Montana by Landsat remote sensing [J]. *Remote Sensing of Environment*, 2009, 113:1843-1852.
- [45] Na X, Zhang S, Li X, et al. Improved land cover mapping using random forests combined with Landsat Thematic Mapper imagery and ancillary geographic data [J]. *Photogrammetric Engineering and Remote Sensing*, 2010, 76:833-840.
- [46] Sesnie S E, Gessler P E, Finegan B, et al. Integrating Landsat TM and SRTM-DEM derived variables with decision trees for habitat classification and change detection in complex neotropical environments [J]. *Remote Sensing of Environment*, 2008, 112:2145-2159.
- [47] Ghimire B, Rogan J, Miller J. Contextual land-cover classification: incorporating spatial dependence in land-cover classification models using random forests and the Getis statistic [J]. *Remote Sensing Letters*, 2010, 1:45-54.

- [48] Prasad A M, Iverson L R, Liaw A. Newer classification and regression tree techniques: bagging and random forests for ecological prediction [J]. *Ecosystems*, 2006, 9:181-199.
- [49] Maclin R, Opitz D. Popular ensemble methods: an empirical study [J]. *Journal of Artificial Intelligence Research*, 1999, 11:169-198.
- [50] Melgani F, Bruzzone L. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 2004, 42:1778-1790.
- [51] Trimble. eCognition Developer 8.7 - Reference Book. Trimble Documentation, 2011.
- [52] 王文字, 李博. 基于 eCognition 的高分辨率遥感图像的自动识别分类技术. 北京建筑工程学院学报. 2006, 22(4):26-29.
- [53] 李弼程, 邵美珍, 黄洁. 模式识别原理与应用[M]. 西安电子科技大学出版社, 2008.
- [54] Koller D, Sahami M. Toward optimal feature selection [J]. In: Saitta L, Eds. *Proceedings of the Thirteenth International Conference on Machine Learning*. San Fransisco: Morgan.
- [55] John G, Kohavi R, Pfleger K. Irrelevant features and the subset selection problem [C]. *The Eleventh Int'l Conf. Machine Learning*, 1994:121-129.
- [56] Molina L C, Belanche L. Feature Selection Algorithm: a Survey and Experimental Evaluation, In *Proc. 2002 IEEE International Conference on Data minging*, 9-12 Dec.2002, 306-313.
- [57] 孙继祥. 现代模式识别[M]. 长沙: 国防科学出版社, 2002.
- [58] 王娟, 慈林林, 姚康泽. 特征选择方法综述[J]. *计算机工程与科学*, 2005, 27(12):68-71.
- [59] 郑毅, 武法东, 刘艳芳. 一种面向对象分类的特征分析方法[J]. *地理与地理信息科学*, 2010, 26(2):16-22.
- [60] 李敏, 崔世勇. 面向对象的高分辨率遥感影像信息提取——以耕地提取为例[J]. *遥感信息*, 2008, (6):64-65.
- [61] 陈忠, 赵忠明. 基于区域生长的多尺度遥感图像分割算法[J]. *计算机工程与应用*, 2005, 41(35):7-8.
- [62] Espindola G M, Camara G, Reis I A, et al. Parameter selection for region-growing image segmentation algorithms using spatial autocorrelation [J]. *Remote Sensing*. 2006, 27(14): 3035-3040.
- [63] Lerman R I, Yitzhaki S. A note on the calculation and interpretation of the Gini index[J]. *Economics Letters*, 1984, 15(3-4):363-368.
- [64] Menze B H, Kelm B M, Masuch R, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data [J]. *BMC Bioinformatics*, 2009, 10(213).
- [65] Strobl C, Boulesteix A L, Zelileis A, et al. Bias in random forest variable importance measures: Illustrations, sources and a solution [J]. *BMC Bioinformatics*, 2007, 8(25).
- [66] R-Development-Core-Team (2009). R: A language and environment for statistical computing -Version 2.10.0. Vienna, Austria: R Foundation for Statistical Computing.
- [67] Liaw, A. (2010). randomForest: Breiman and Cutler's random forests for classification and regression, Version 4.5-36.

-
- [68] Kumar K V, Martha T R, Roy P S. Mapping damage in the Jammu and Kashmir caused by 8 October 2005 M-w 7.3 earthquake from the Cartosat-1 and Resourcesat-1 imagery [J]. International Journal of Remote Sensing, 2006, 27(20):4449-4459.
- [69] Martha T R, Kerle N, van Westen C J, et al. Effect of Sun Elevation Angle on DSMs Derived from Cartosat-1 Data [J]. Photogrammetric Engineering and Remote Sensing, 2010, 76(4):429-438.
- [70] Cruden D M, Varnes D J. Landslide types and processes [C]. Turner A K, Schuster R L. Landslides: Investigation and Mitigation: Special Report 247. Washington D.C: Transportation Research Board, 1996:36-75.
- [71] 陈永刚, 汤国安, 祝士杰等. DEM 重采样误差空间分布格局及差异性分析[J]. 中国矿业大学学报, 2011, 40(4):653-659.
- [72] 王宇, 白天路. 重采样方法对 DEM 地形表达的影响分析[J]. 工程勘察, 2013, 3:67-70.
- [73] Strahler A N. Introduction to Physical Geography[C]. Wiley & Sons, New York, 1965.