

利用监督分类快速提取分层设色图件中数据的方法

刘 晓¹, 黄海军², 孙荣涛¹, 张照录¹, 宋明春³, 董慧君⁴

(1. 山东理工大学 资源与环境工程学院, 山东 淄博 255049; 2. 中国科学院 海洋研究所, 海洋地质与环境重点实验室, 山东 青岛 266071; 3. 山东省地质矿产勘查开发局, 山东 济南 250013; 4. 东营市勘察测绘院, 山东 东营 257091)

摘要: 为了快速将分层设色图件中象元的颜色值转化为具有实际意义的属性标识或物理、化学参数值, 从而提取分层设色图件中包含的数据, 作者提出了根据不同颜色分量(RGB)构建颜色特征空间, 基于最小距离监督分类方法快速提取分层设色图件中数据的方法。使用该方法对东半球海域海面水汽分布数据进行提取, 取得了良好的效果。该方法可以广泛应用于物理海洋、环境科学、地质调查、地理信息系统建库等多个行业领域的的数据提取工作中。

关键词: 监督分类; 数据提取; 分层设色法; ENVI

中图分类号: P208.2; P285.3 文献标识码: A 文章编号: 1000-3096(2016)07-0110-05

doi: 10.11759/hyxx20141103004

1 研究背景

伴随全球变化研究的开展及国际交流合作水平的提高, 国内外许多机构定期或不定期地发布海洋、气象、环境等领域的监测数据及研究成果^[1-3]。这些监测数据及研究成果对于科学研究、工程建设、环境保护等都有重要意义^[4]。这些数据及成果, 有的以原始数据的形式发布, 有的以论文专著中图件的形式发布, 有的则从建立的数据发布网站发布^[5-7]。对于以图像形式发布的数据, 一般使用象元颜色代表某种属性或参数, 且使用的是图像坐标系。在进行模型设计、分析处理时, 较难利用其中包含的数据信息。如何快速从这些图像中快速提取有用数据, 将图像坐标系下象元颜色值转换为地理坐标系或投影坐标系下属性或物理、化学参数的值, 成为科学研究及工程实践中经常碰到的问题^[8-9]。

对于图像中用等值线法描述的数据, 可以使用地理信息系统软件进行半自动化屏幕跟踪数字化, 然后根据等值线进行内插即可完成数据的提取^[10-11]。这种方法已经非常成熟, 多数地理信息系统软件都有相关功能, 作者不再赘述。另外有一些数据使用分层设色法(Hypsometric law)进行描述, 以某个色系中不同的颜色表示不同的数值大小。这些数据或者来自遥感卫星监测结果, 或者来自于海洋大气模型模拟的结果, 通常使用 Matlab、IDL、Fortran、C 等语言编写的程序进行绘制。这些数据中没有明显的等值

线, 也很难分辨表示某一数据范围的多边形, 数据较为散乱, 很难用手工数字化的方法进行数据提取^[12-13]。本研究提出一种使用监督分类法快速提取分层设色图像中原始数据的方法, 以该方法提取东半球海域海面水汽分布数据, 获得了良好的效果。

2 使用监督分类法快速提取图像中原始数据的原理

分层设色图件通常是由软件生成的, 通常包括反映分层信息的颜色以及辅助要素(如注记、空白、图框)使用的颜色两类。这些图件一般采用栅格图像格式存储, 对于图像中每个像素, 有对应的红、绿、蓝 3 个色彩分量。可以建立三维特征空间, 如图 1 所示, 其 x 、 y 、 z 轴分别对应图像的红、绿、蓝 3 个色彩分量。对于图像中每种分层所对应的颜色, 在该特征空间中为一个特征点。栅格图像中的每个像素, 根

收稿日期: 2014-11-03; 修回日期: 2015-02-28

基金项目: 国家自然科学基金项目(41006021, 41306190); 山东省自然科学基金项目(ZR2014DQ005); 泰山学者工程专项经费资助

[Foundation: National Natural Science Foundation of China, No. 41006021, No. 41306190; Natural Science Foundation of Shandong Province, China, No.ZR2014DQ005; Taishan Scholars Program of Shandong Province]

作者简介: 刘晓(1981-), 男, 山东淄博人, 博士, 副教授, 主要从事现代海洋沉积环境研究, 电话: 15169225534, E-mail: liuxiaogis@sdu.edu.cn; 黄海军, 通信作者, 研究员, 电话: 053282898524, E-mail: hjhuang@qdio.ac.cn

据其 R、G、B 值都能在该特征空间中找到对应的位置。计算该像元位置与每个分层对应的特征点之间的距离,该位置最靠近的特征点对应的分层,即是该像元所属的分层^[14]。对图像中每个像元进行上述判别操作,可以快提取图像中各个分层的范围。该操作在遥感图像数字处理中称为最小距离监督分类法(supervised classification),被广泛用于遥感数据分类^[15-16]。

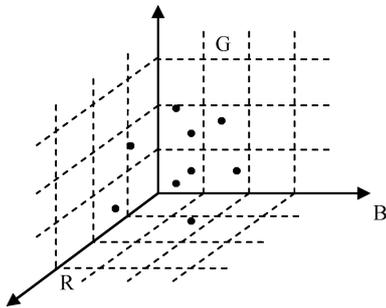


图 1 根据 RGB 波段建立三维特征空间

Fig. 1 Feature space according to the RGB band

3 数据来源及处理流程

3.1 图片来源

本研究使用的图像来自 OSPO(The Office of Satellite and Product Operations)网站(<http://www.ospo.noaa.gov/>)。OSPO 是隶属与 NOAA 的一个部门,主要负责卫星数据的处理及信息的发布。OSPO 网站发布的数据中,包括陆地、海洋、大气多种产品。本研究中使用的是美国空军,美国海军,美国海洋大气总署共同参与了 SPP(Shared Processing Program 共享处理程序)项目中得到的东半球海域海面水汽分布图,进行东半球海面水汽分布数据的提取。本研究中使用的原始图像自 http://www.ospo.noaa.gov/data/spp/ssmi/water_vapor.east.gif 网址下载,如图 2 所示。

3.2 数据处理流程

ENVI (The Environment for Visualizing Images, ENVI)是由遥感领域的科学家采用交互式数据语言 IDL(Interactive Data Language, IDL)开发的一套功能强大的遥感图像处理软件,其功能覆盖了图像数据处理的几乎所有功能^[15]。ENVI 不但允许用户使用 IDL 语言对其特性和功能进行扩展或自定义,甚至将 C++和 Java 代码集成到其工具中。本研究在 ENVI4.8 软件中使用监督分类法快速提取图片中原始数据,其数据处理流程如图 3 所示。

3.2.1 导入数据并采集控制点

由于 ENVI 软件不支持直接打开 GIF 格式的图像,故首先将网站得到的 GIF 格式图像保存为无压缩的 BMP 格式的图像文件,在 ENVI 软件中打开。注意不要保存为 JPG 格式,该格式对图像进行有损压缩过程中,会改变某些像元的颜色,影响分类精度。对其坐标系统进行分析发现,该图中经纬线皆为直线且相互垂直,相同间隔的经纬线组成正方形网格,故判断其使用经纬度坐标系,其横纵坐标轴为地理经纬度。横坐标范围从东经 20°向东增大到 180°,纵坐标从南纬-80°向北越过赤道增加到北纬 80°。根据以上信息,在 ENVI 中使用“选择地面控制点:图像配准到地图(Select GCPs: Image to Map)”工具采集配准使用的控制点。由于配准过程中会对原始图像进行重采样,进而影响监督分类的精度,故在此步骤中不进行纠正操作,仅采集控制点并保存在文件中。

3.2.2 图面内容分析

图中使用分层设色法表示海面水汽分布情况,共划分为 20 个等级,除此之外还有陆地、注记及经纬线、无数据区、图像背景处对应的颜色。使用“光标位置/值(Cursor Location / Value)”工具查看发现其中图像背景、陆地及水汽浓度小于 3.5 处像元的 RGB 值相同,其他各类像元 RGB 值各不相同。根据本次试验数据,在监督分类中共建立了 22 类样本。

3.2.3 选择监督分类样本

在“感兴趣区工具(ROI Tool)”中建立样本类,并在显示窗中采集样本。由于图例是表示不同等级水汽含量的标准,且图例大小合适,类型单一,较少受到其他像元的干扰,故在采集样本时优先选择在图例处进行采集。采集样本过程中,在“感兴趣区工具(ROI Tool)”中将每类样本的名字更改为能反映含义的文字或数字,并使用“保存感兴趣区(Save ROIs)”工具保存为*.roi 文件,以方便后期处理。

3.2.4 最小距离法进行监督分类

完成分类样本采集后,即可使用最小距离法进行监督分类。分类结果存储在与原图像同样大小的图像文件中,其栅格代码及颜色对应于建立的样本文件。对比原图与分类后的结果,发现分类效果非常好,主要原因是图像中色彩类型较单一,仅在不同颜色交界处有少数像元颜色受到周围像元的影响。

3.2.5 数据后处理

由于原图像是由软件对数据不同范围赋以特定颜色生成的,故原图像中不存在遥感解译种种常见

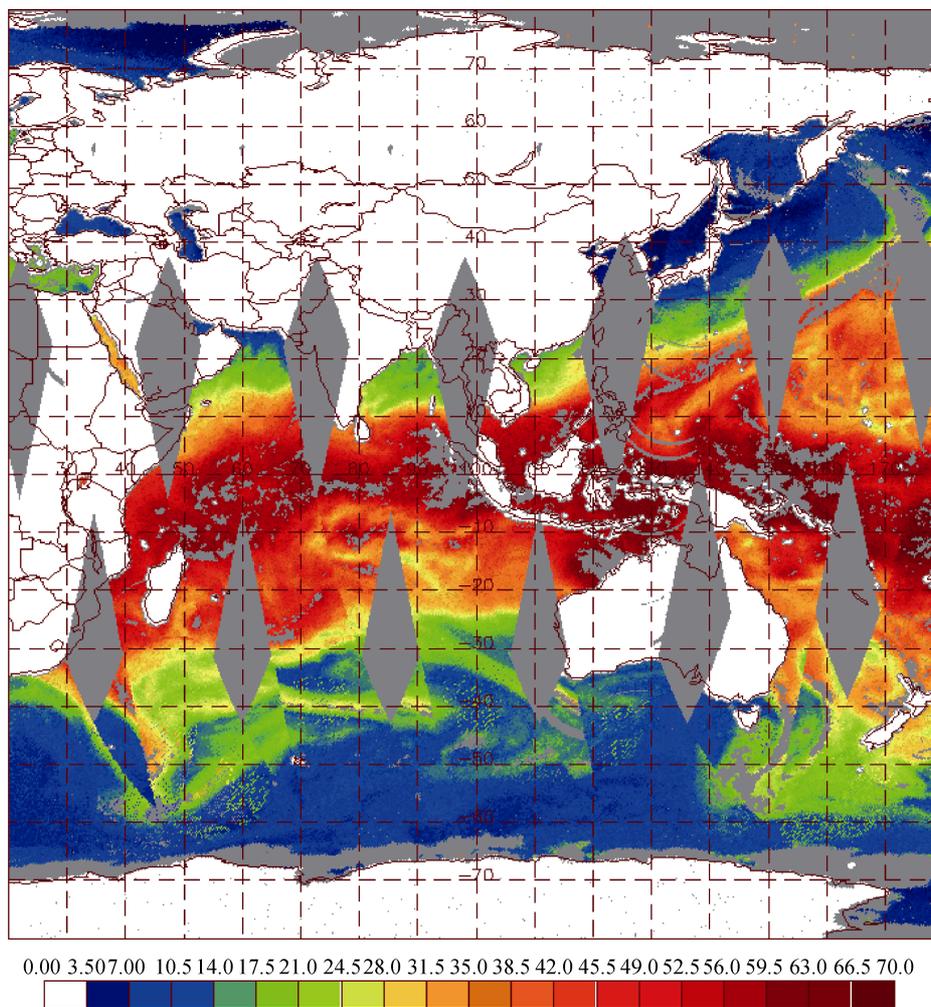


图 2 东半球海域海面水汽分布图(UTC: 2013 年 12 月 28 日 20 点)

Fig. 2 Map of the water vapor above the sea in the eastern hemisphere

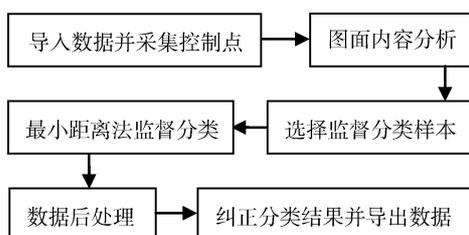


图 3 使用监督分类法提取分层设色图片中数据的流程
Fig. 3 Process used to extract data from hypsometric law-based map with supervised classification method

的同物异色、混合像元、地形干扰等问题,故不需要进行过多的后处理操作。确实有需要,可以通过综合运用“类别集群(Clump)”、“类别筛选(Sieve)”工具消除分类结果中的小斑块。

3.2.6 纠正分类结果并导出数据

根据第一步保存的控制点文件,使用“基于地面控制点纠正(Wrap from GCPs: Image to Map)”工

具对分类结果图像进行纠正。使用“根据感兴趣区裁切(Subset Data via ROIs)”工具裁切掉图名、图例、比例尺、图框等内容,仅保留有数据的区域。将数据另存为其他格式(如*.img),完成数据提取工作。

3.2.7 提取结果精度评价

将上述操作提取到的结果数据保存为 data.img 文件并导入 ArcGIS 软件。使用相同的控制点文件对未分类的原始图像进行纠正,将得到的数据另存为 source.img。在 ArcGIS 软件中,使用“创建随机点(Creat Random Points)”工具,创建随机点用以检验数据提取效果。根据本研究的实际情况,此处创建了 300 个随机点进行检验。使用“值提取至点(Extract Multi Value to Points)”工具,分别提取随机点所在位置处 data.img、source.img 的 RGB 图层的像元值,并进行统计以判断该点分类是否准确。

4 结果与结论

将本研究中提取到的数据在 ArcGIS 软件中制图, 结果如下图 4 所示^[10]。该图与原图对比进行目视检查可以看出, 数据提取效果非常好。在本次试验中通过统计得出, 类别划分正确的象元数占总象元数的 92.3%。类别划分错误主要来源于原始图像中经纬线的干扰, 占总错误数的 69.6%, 这部分错误对于一般的建模及统计分析没有影响。其他操作导致的错误, 仅占总错误数的 30.4%, 占总象元数的 2.33%。该提取结果完全能够满足一般工程应用及科学研究的需要。

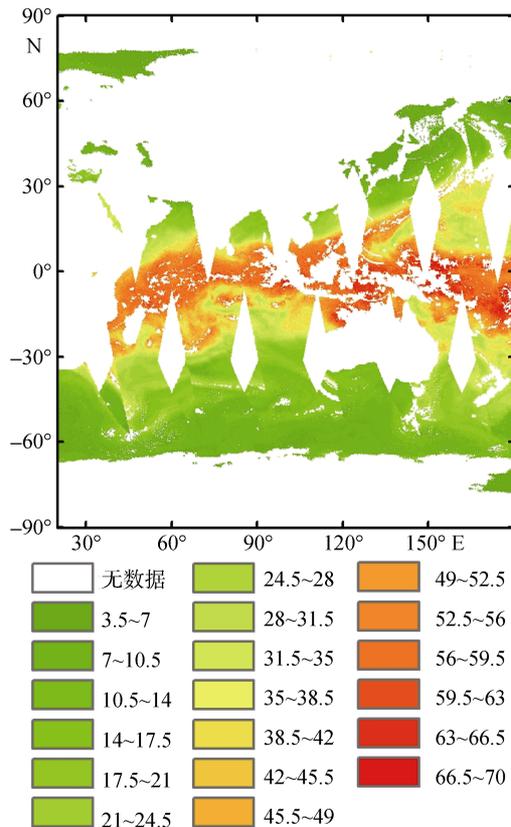


图 4 使用提取的数据绘制的东半球海域海面水汽分布图
Fig. 4 Distribution map of the water vapor above the sea in the eastern hemisphere according to data extracted above

将提取得到的结果与原始图件进行比较, 发现误差产生的原因有以下几点: (1)经纬线、图像边框及数字标注线产生的干扰; (2)小斑块消除运算过程中产生的误差; (3)图像保存为 BMP 格式及图像配准过程中对原始图像进行重采样引入的误差; (4)小斑块合并带来的误差。

上述误差仅在非常少量的局部地区对数据提取

结果产生影响, 从整体来看, 本研究中提出的方法快捷方便, 能快速准确的从原始分层设色图像中提取连续变化的数据信息。

本研究中根据监督分类法原理, 建立基于 RGB 颜色分量的特征空间, 根据图像中不同类型要素对应的颜色, 使用最小距离法进行分类, 快速准确地提取了分层设色图像中的原始数据。该方法提取数据操作简便、速度快、准确性高, 解决了过去使用手工方法难以提取该种类型数据的问题。该方法具有广阔的应用前景, 可以广泛应用于物理海洋、环境科学、地质调查、地理信息系统建库等多个行业领域。

参考文献:

- [1] 袁雪梅, 蒋永国, 郭忠文. 海洋数据信息共享平台关键技术研究与实现[J]. 中国海洋大学学报(自然科学版), 2010, 40(12): 147-153.
Yuan Xuemei, Jiang Yongguo, Guo Zhongwen. Design and implementation of marine data information sharing platform[J]. Periodical of Ocean University of China, 2010, 40(12): 147-153.
- [2] 徐超, 李莎, 米浦春. 南海物理海洋数据的 OPeNDAP 服务实现[J]. 热带海洋学报, 2010, 29(4): 174-180.
Xu Chao, Li Sha, Mi Puchun. OPeNDAP service based implementation of physical oceanographic data of the South China Sea[J]. Journal of Tropical Oceanography, 2010, 29(4): 174-180.
- [3] 刘海行, 宋转玲, 韩彬. 青岛海洋资料共享服务系统的设计与实现[J]. 海洋科学进展, 2008, 26(3): 394-400.
Liu Haixing, Song Zhuanling, Han Bin. The design and realization of Qingdao oceanic data-sharing service system[J]. Advances in Marine Science, 2008, 26(3): 394-400.
- [4] 杨峰, 杜云燕, 崔马军, 等. 海洋多源环境数据共享系统研究[J]. 测绘通报, 2011, 10: 35-37.
Yang Feng, Du Yunyan, Cui Majun, et al. Study of multi-source marine environmental data sharing system[J]. Bulletin of Surveying and Mapping, 2011, 10: 35-37.
- [5] 宋德瑞, 陶冠峰, 孙淑艳, 等. 海洋环境监测业务共享平台建设与实践[J]. 海洋环境科学, 2008, 27(S2): 114-116.
Song Derui, Tao Guanfeng, Sun Shuyan, et al. Construction and implement of business share platform in marine environmental monitoring[J]. Marine Environmental Science, 2008, 27(S2): 114-116.
- [6] Zhang Y, Wang J, Vorontsov A M, et al. Using a neural network approach and time series data from an international monitoring station in the Yellow Sea for model-

- ing marine ecosystems[J]. *Environ Monit Assess*, 2014, 186(1): 515-524.
- [7] O'Hara P D, Serra-Sogas N, Canessa R, et al. Estimating discharge rates of oily wastes and deterrence based on aerial surveillance data collected in western Canadian marine waters[J]. *Mar Pollut Bull*, 2013, 69(1-2): 157-164.
- [8] 薛存金, 苏奋振, 杜云艳. 海洋地理信息系统集成技术分析[J]. *海洋学报(中文版)*, 2008, 30(4): 56-62.
Xue Cunjin, Su Fenzhen, Du Yunyan. Analysis of integration technology based on marine geographical information system[J]. *Acta Oceanologica Sinica*, 2008, 30(4): 56-62.
- [9] 赵冬泉, 党安荣, 陈吉宁. 监督分类方法在图片资料专题信息提取中的应用研究[J]. *测绘通报*, 2006, 11: 32-34.
Zhao Dongquan, Dang Anrong, Chen Jining. A Study of application of supervised classification to thematic information extraction in image data[J]. *Bulletin of Surveying and Mapping*, 2006, 11: 32-34.
- [10] 汤国安, 杨昕. ArcGIS 地理信息系统空间分析实验教程[M]. 北京: 科学出版社, 2012: 289-296.
Tang Guoan, Yang Xin. ArcGIS geography information system spatial analysis experiment instructor[M]. Beijing: Science Press, 2012: 289-296.
- [11] Hoenner X, Whiting S D, Hindell M A, et al. Enhancing the use of Argos satellite data for home range and long distance migration studies of marine animals[J]. *PLoS One*, 2012, 7(7): e40713.
- [12] Giannakeas N, Fotiadis D I. Multichannel segmentation of cDNA microarray images using the Bayes classifier[J]. *Conf Proc IEEE Eng Med Biol Soc*, 2007: 3466-3469.
- [13] Claeys A, Cornelis A, Kerckaert I, et al. Fully automated measurements by light microscopy of tissue sections using a cellular array computer[J]. *Gegenbaurs Morphol Jahrb*, 1989, 135(1): 83-90.
- [14] Baattrup-Pedersen A, Larsen S E, Riis T. From expert judgement to supervised classification: a new approach to assess ecological status in lowland streams[J]. *Sci Total Environ*, 2013, 447: 116-122.
- [15] 邓书斌. ENVI 遥感图像处理方法[M]. 北京: 科学出版社, 2010: 129-145.
Deng Shubin. ENVI remote sensing image process method[M]. Beijing: Science Press, 2010: 129-145.
- [16] 韩丽娟, 王鹏新, 王锦地, 等. 植被指数-地表温度构成的特征空间研究[J]. *中国科学(D 辑: 地球科学)*, 2005, 35(4): 371-377.
Han Lijuan, Wang Pengxin, Wang Jindi, et al. Research on surface temperature/NDVI index feature space[J]. *Science In China Ser(D Earth Sciences)*, 2005, 35(4): 371-377.

Application of supervised classification method in the extraction of data from hypsometric law-based map

LIU Xiao¹, HUANG Hai-jun², SUN Rong-tao¹, ZHANG Zhao-lu¹, SONG Ming-chun³, DONG Hui-jun⁴

(1. College of Resource and Environmental Engineering, Shandong University of Technology, Zibo 255049, China; 2. Key Lab of Marine Geology and Environment, Institute of Oenology, Chinese Academy of Sciences, Qingdao 266071, China; 3. Exploration and Development Bureau of Geology and Mineral Resources of Shandong Province, Jinan 255013, China; 4. Dongying Exploration and Surveying Institute, Dongying 257091, China)

Received: Nov. 3, 2014

Key words: Supervised classification method; data extraction; hypsometric law-based map; ENVI

Abstract: In this article, a new method was designed to convert the RGB value of pixels in the hypsometric law-based map to its real property code, to convert it quickly to its physical/chemical parameters, and to extract spatial data from the hypsometric law-based map. The minimum distance supervised classification method was used to construct the feature class space based on the RGB component of the map image. This method also provided a good result when extracting water vapor concentration data in the Eastern Hemisphere and can thus be used widely in data extraction work in fields such as physical oceanology, environmental science, geological surveying, and GIS database construction.

(本文编辑: 谭雪静)