

居民地信息提取的最优波段组合方案研究

——以开封市区为例

冯恩国,秦 奋,李淑香

(河南大学黄河文明与可持续发展研究中心,河南大学环境与规划学院,河南 开封 475001)

摘 要:遥感以其高光谱分辨率、空间分辨率为地理研究提供了丰富的数据,这些数据各波段之间具有重要的互补性,但是也存在着大量的数据冗余。利用最佳波段组合指数法和地物光谱信息相结合的方法,按照波段组合后的信息量最大和波段之间的相关性最小的原则,对实验区的居民地信息提取选取最佳波段组合方案,最后研究表明:分别将 TM4、TM5、TM3 波段赋予红绿蓝三种颜色,所得的假彩色图像信息丰富,居民地与周围地物在色调上具有很好的可分性,有利于信息的提取。

关键词:居民地信息;最优波段组合;OIF 指数

中图分类号: TP75

文献标识码: A

文章编号: 1005-3409(2007)03-0200-03

Study on Optimal Bands Composite in Extracting Residential Areas Information

FENG En-guo, QIN Fen, LI Shu-xiang

(Key Research Institute of Yellow River Civilization and Sustainable Development, Henan University, College of Planning and Environment, Henan University, Kaifeng 475001, China)

Abstract: The remote sensing provide geographical research rich data with its high spectrum resolution and the spatial resolution. These data have very important supplementary, but also have massive data redundancy. The authors use the OIF method and consider the spectrum information of target objects. In this study, the follow factors must be consider: the information of wave bands and the difference of target objects' spectrum information must be max, the correlation of the wave bands must be lower. In the end, this study indicate: the false color image composed of TM4, TM5, TM3 wave bands is rich in information and the residential areas can obviously differ from surrounding objects in color. So it is better for extracting the information of residential areas.

Key words: information of residential areas; optimal bands composite; OIF index

遥感技术自上世纪 60 年代发展以来以其感知面积大、数据获取周期短、信息量大等特点为地理研究提供了大量数据。20 世纪末,遥感技术迅猛发展,携带光学、微波、多光谱、高光谱等各种功能的传感器的卫星陆续升空,为人类进行资源、环境、灾害与监测更是提供了海量数据^[1,2]。它们以不同的光谱分辨率,反映地物目标在各个光谱范围内的各种特性,形成了同一地区的多光谱遥感数据。应用多光谱遥感数据可以得到大量的地物光谱特征,但是目前在进行影像判读与专题信息提取时仍离不开目视解译。由于人眼对彩色敏感且分辨力强,多光谱数据往往需要进行波段组合合成假彩色图像以便于进行目视解译和专题信息提取。由于多光谱遥感数据波段较多,且各波段之间既具有互补性,又存在着大量的数据冗余,如何对多光谱数据进行最佳波段组合以便于快速、准确提取目标地物信息成为遥感图像处理的关键之一。

1 波段组合原理

遥感技术发展到现在,目视解译在遥感信息提取中仍占重要的位置,在目视判读时一般都是选用信息丰富的彩色

合成图像。假彩色图像是根据加色法彩色合成方案,选择多光谱影像中的某三个波段,分别赋予红、绿、蓝三原色组合而成的^[3]。多光谱遥感数据波段较多,如何进行方案组合是假彩色合成的关键。波段选择要遵循以下三个原则:(1)所选的三个波段的信息总量要大;(2)所选的三个波段的相关性要弱;(3)目标地物类型要在所选的波段组合内与其它地物具有很好的可分性^[4]。

依据以上三个原则,在波段选择时主要依据是:方差越大,表明波段的离散程度越大,所含的信息量越丰富,因而在进行波段组合时首先要考虑组合后的波段方差要尽可能的大。考虑到第二个原则,由于地物各个波段之间的辐射特性存在相关性,三个方差最大的波段组合在一起不一定能得到最大的信息量。此外,在波段选择时仅仅考虑上面两个原则是不够的,当满足上面两个原则,但所研究的地物在所选的三个波段与其它地物不具有很好的可分性,波段组合仍不是最佳的波段组合。因此,在进行假彩色方案合成时,波段的选取一定要考虑方差要大而相关性要小这样两个条件,在此基础上目标地物还应该与其它地物在此波段组合上具有很好的可分性。

* 收稿日期:2006-06-01

基金项目:国家科技基础条件平台建设项目(2005DKA32300-08);水利部科技创新基金项目(SCX2002-08);河南省社会科学基金项目(2003FSH006)

作者简介:冯恩国(1980-),男,山东聊城人,河南大学环境与规划学院硕士研究生,主要从事 GIS、RS 的应用研究。

2 研究区概况及数据分析

本文以开封市区为实验区,寻求居民点信息提取的最佳波段组合方案。

2.1 实验区概况

开封市位于北纬 34°11'43"~35°01'33",东经 113°51'57"~115°15'42"之间,南北宽约 92 km,东西长约 126 km,土地总面

积 6 282.40 km²。现辖开封、兰考、杞县、通许、尉氏县和龙亭、鼓楼、顺河、南关、郊区,共五县五区。本研究只选取开封市区及郊区部分地区。其经纬度范围为北纬 34°40'31"~34°55'26",东经 114°11'36"~114°30'01"。实验区位于豫东冲击平原东部,黄河冲积扇上,总的趋势是由西北向东南倾斜,海拔高度介于 58~78 m。地势平坦,居民地密集分布,且受地形变化影响较小。研究区的大致位置如图 1 所示。

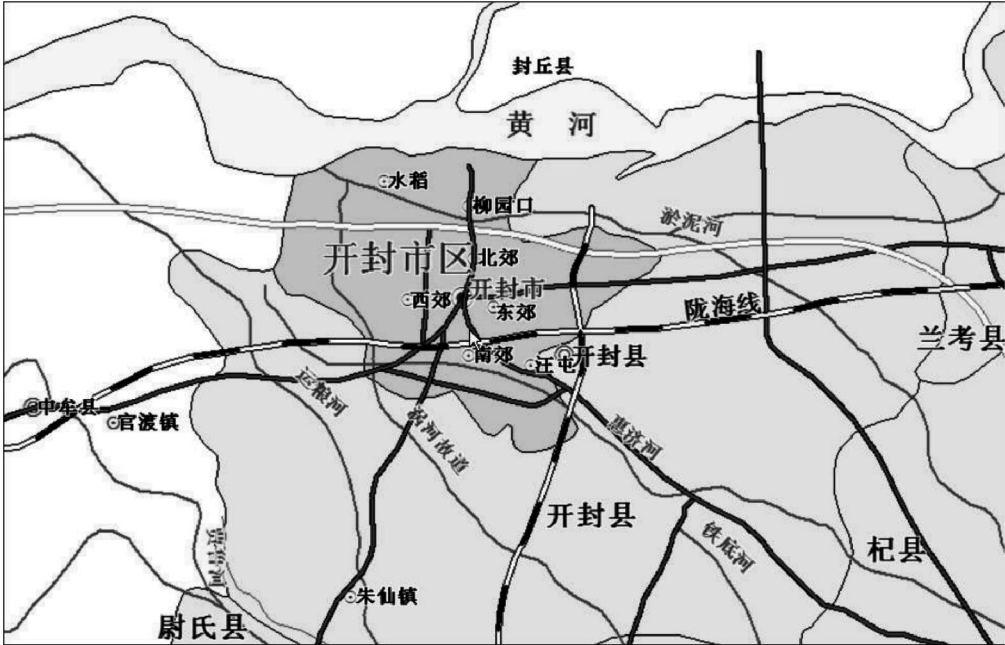


图 1 研究区范围示意图

2.2 多光谱数据

美国陆地卫星系列的 TM 影像因其较高的空间分辨率和时间分辨率已经广泛的应用于土地利用、农林业、生态等领域。TM 影像数据的观测参数如表 1:

表 1 TM 的观测参数^[5]

波段	波长/ μm	空间分辨率/m
1	0.45~0.52 蓝色	30
2	0.52~0.60 绿色	30
3	0.63~0.69 红色	30
4	0.76~0.90 近红外	30
5	1.55~1.75 短波红外	30
6	10.40~12.50 热红外	120

本次实验所采用的遥感影像数据是采集于 2002 年 9 月 2 日

TM7 号星的多光谱数据。其轨道号为 124 - 36。开封市位于该景图像上的右上角。根据在 1:5 万的地形图上确定的研究区范围在 ERDAS 软件中对该景影像进行裁减处理,获得研究区的影像图。影像图如图 2 所示,波段组合为标准假彩色 432 波段组合。根据表 1,影像的 1 - 5,7 六个波段的空间分辨率为 30 m,第 6 波段其空间分辨率为 120 m,为红外波段,主要反映地物的温度信息,与其它六波段相比,分辨率较低。因此,在波段组合研究中只选取了 1~5 和 7 波段。

2.3 实验区各地物光谱特征分析

遥感图像本身包含大量的地物波谱信息,利用矢量化后的开封市土地利用现状图与遥感影像叠合,选取各土地利用类型的若干样本,获取遥感影像同步的地物光谱信息。利用这些统计数据可以做出各地物的光谱曲线图(图 4)。

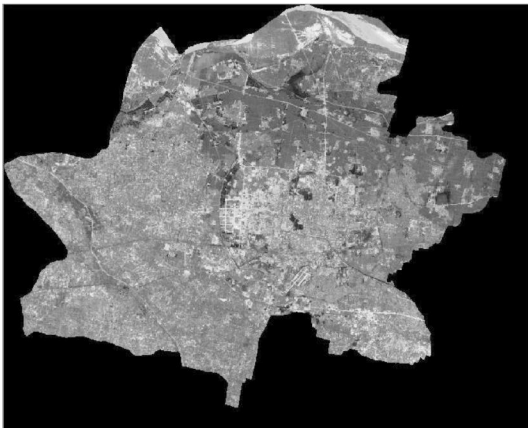


图 2 研究区影像裁减图(Band432)

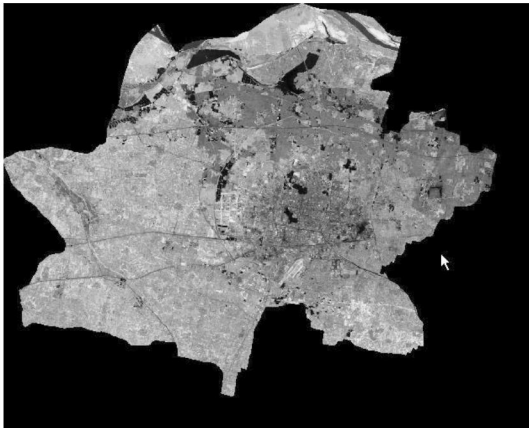


图 3 453 波段组合假彩色图像

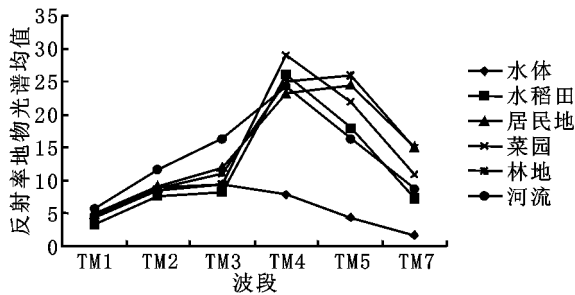


图 4 TM 影像各地物反射率曲线图

从图中可以看出在 TM1 和 TM2 波段,居民地与其它地物差异不大,目标地物很难与背景地物区分开来。从 TM3 波段开始居民地与其它地物的差异开始逐渐增大,在 TM5 波段时达到各地物的差异最大。

3 最佳波段选择

目前,最佳波段选择的方法有很多种^[5]:最佳波段组合指数法(Optimum Index Factor ,OIF 法)、方差 - 协方差矩阵特征值法、主成分分析法、熵与联合熵法等。本文采用简便实用的 OIF 指数法,进行最佳波段组合研究。本次研究是在 PC 机上进行的,所采用的影像处理环境包括:ERDAS IMAGE8.5 和 ENVI3.6。

3.1 最佳波段组合指数法(OIF 指数法)的计算

该方法是美国的 Chavez 教授提出的,它的基本思想是:利用 3 个波段的标准差及两两之间的相关系数计算一个最佳指数因子(optimum index factor ,OIF),图像数据的标准差越大,所包含的信息量也越多,而波段的相关系数越小,表明各波段的图像数据独立性也就越高,信息的冗余度也就越小^[6]。其数学表达式如下:

$$OIF = \frac{\sum_{i=1}^3 S_i / \sum_{j=1}^3 |r_{ij}|}$$

式中: S_i ——第 i 波段的标准差, r_{ij} ——第 i 波段与第 j 波段之间的相关系数。 OIF 值越大,说明此 3 个波段组合所包含的信息量越大,波段的相关性越小。

在 OIF 指数的计算中,关键是求出各个波段的标准差和波段两两之间的相关系数。相关系数可由相关系数公式求出:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}$$

式中: $i, j = 1, 2, 3, 4, 5, 7$ ——波段数, x_{ki} ——第 i 波段第 k 个像元的灰度值, \bar{x}_i ——第 i 波段的像元灰度均值, x_{kj} ——第 j 波段的第 k 像元的灰度值, \bar{x}_j ——第 j 波段的像元灰度均值。 $K = 1, 2, 3 \dots, n$ ——实验区的像元数。

标准差可以在 ERDAS IMAGE 8.5 中的图层信息中求得。相关系数在 ERDAS IMAGE 8.5 中没有相应的模块可以直接计算出来,不过可以通过 ERDAS IMAGE 中的 MODLE 模块开发相应的计算模型计算出各波段的相关系数,但是过程相对烦琐。本文通过另一款遥感图像处理系统 ENVI3.6 来计算各波段的相关系数。在 ENVI3.6 中的 BASIC TOOLS 的 STATISTICS 模块中,设置计算影像的统计信息,可以统计出图像的标准差和各波段的相关系数矩阵。各波段的标准差及相关系数矩阵表如表 2:

将表 2 中各项结果带入 OIF 指数计算公式,得到表 3:

3.2 结果分析与最佳波段组合

从表 2 结果分析, TM1, TM2 与其它波段的相关性较

大,尤其是 TM1 和 TM2, TM2 和 TM3 之间,分别为 0.995 533 和 0.984 495。在后面的波段组合中应该尽量避免使用 TM1 和 TM2 波段。TM5 波段与其它波段之间的相关性较小,是一个相对独立的波段。从表 3 结果分析,先不考虑其它因素,仅就 OIF 值来说,波段组合 145, 345, 457, 135, 245 的 OIF 值较高,是最佳波段组合备选组合。结合本实验的研究目的是为了提取居民地信息,要结合当时的地物光谱信息来选择最佳波段。在蓝色波段(TM1)和绿色波段(TM2),居民地除与河流能很好的分离,与其它背景地物可分性较差,到红色波段,地物之间的差异开始加大。考虑到由表 2 分析后的结果,在波段选择时不选择 TM1、TM2 波段参与组合。排除波段组合 145, 135, 245。按照 OIF 值的大小决定 345 为最佳波段组合。将 TM3、TM4、TM5 分别赋予红绿蓝三原色又有六种组合方案,通过目视分辨,将 TM4 赋予红色、TM5 赋予绿色、TM3 赋予蓝色(即 453 波段组合),合成后的假彩色影像(如图 3),地物光谱信息丰富,可分性较好。在图像上蓝色、砖红色、青绿色、深蓝色构成了影像一级波谱信息,分别表示居民地、水田、旱地和水体。居民地可以很好的从背景地物中分别出来。

表 2 各波段间的标准差及相关系数矩阵表

波段	band1	band2	band3	band4	band5	band7
band1	1					
band2	0.995533	1				
band3	0.968132	0.984495	1			
band4	0.921283	0.901289	0.829419	1		
band5	0.945347	0.949657	0.933111	0.911475	1	
band7	0.922361	0.938594	0.955383	0.817618	0.97499	1
标准差	42.0206	35.71535	36.38031	38.91263	48.07985	34.415

表 3 OIF 指数及其排序

波段组合	标准差求和	相关系数求和	OIF 指数	最佳组合排序
123	114.116259	2.94816	38.7076	19
124	116.648583	2.818105	41.3926	13
125	125.815806	2.890537	43.5268	7
127	112.150956	2.856488	39.2618	18
134	117.313536	2.718834	43.1485	9
135	126.480759	2.84659	44.4324	4
137	112.815909	2.845876	39.6419	17
145	129.013083	2.778105	46.4392	1
147	115.348233	2.661262	43.3434	8
157	124.515456	2.842698	43.8019	6
234	111.008289	2.715203	40.8839	16
235	120.175512	2.867263	41.913	11
237	106.510662	2.878472	37.0025	20
245	122.707836	2.762421	44.4204	5
247	109.042986	2.657501	41.0322	15
257	118.210209	2.863241	41.2855	14
345	123.372789	2.674005	46.1378	2
347	109.707939	2.60242	42.1561	10
357	118.875162	2.863484	41.5141	12
457	121.407486	2.704083	44.8978	3

4 结 论

遥感影像的最优波段组合方案的确定是图像信息增强的关键技术之一。波段组合的优劣,直接影响遥感影像信息

(下转第 205 页)

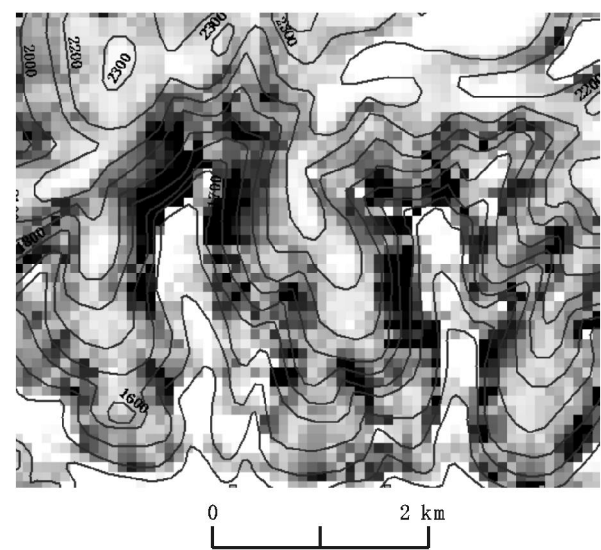


图 3 LS 因子栅格数据与等高线矢量数据叠加图

3 应用实例

在湖北三峡香溪河某个支流流域,用 1:25 万的等高线图数字化后生成 DEM 数据,取栅格单元大小为 30 m×30 m,采用本文所介绍的方法计算通用土壤流失方程中的土壤侵蚀地形因子值,得到最小值为 0,最大值为 190,平均值为

参考文献:

[1] Wischmeier W H, D D Smith. Predicting rainfall erosion losses from cropland east of the rocky mountains: guide for selection of practices for soil and water conservation[Z]. U. S. Dep. Agric., Agric. Handb. No.282,1965.

[2] Wischmeier W H, D D Smith. Predicting rainfall erosion losses: A guide to conservation planning with Universal Soil Loss Equation (USLE)[Z]. U. S. Dep. Agric., Agric. Handb. No.537,1978.

[3] Renard, K G, Foster, G R, Weesies, G A,et al. Predicting soil erosion by water: a guide to conservation planning with Revised Universal Soil Loss Equation (RUSLE) [Z]. U. S. Dep. Agric., Agric. Handb. No.703,1996.

[4] 张照录,薛重生.通用土壤流失方程的研究进展及其改进方法[J].地学前缘,2004,11(4):608.

[5] 蔡崇法,丁树文,史志华,等.应用 USLE 模型与地理信息系统 IDRISI 预测小流域土壤侵蚀量的研究[J].水土保持学报,2000,14(2):19-24.

[6] 肖寒,欧阳志云,王效科,等. GIS 支持下的海南岛土壤侵蚀空间分布特征[J].土壤侵蚀与水土保持学报,1999,5(4):75-80.

[7] 游松财,李文卿. GIS 支持下的土壤侵蚀量估算[J].自然资源学报,1999,14(1):62-67.

[8] Dunn, M, R Hickey. The effect of slope algorithms on slope estimates within a GIS[J]. Cartography, 1998, 27(1):9-15.

[9] Hickey, R A Smith, P Jankowski. Slope length calculations from a DEM within ARC/INFO GRID[J]. Computers, Environment and Urban Systems, 1994, 18(5):365-380.

(上接第 202 页)

的多少,进一步影响后面影像的目视解译和专题信息提取。本文在实验区内,利用最佳波段组合指数法(OIF 法)与地物光谱信息相结合,最后确定 453 波段为最佳波段组合,在该

参考文献:

[1] 梅安新,彭望碌,等.遥感导论[M].北京:高等教育出版社,2001.

[2] 赵英时,等.遥感应用分析原理与方法[M].北京:科学出版社,2003.

[3] 戴昌达,姜小光,等.遥感图像应用处理与分析[M].北京:清华大学出版社,2004.

[4] 姜小光,王长耀等.成像光谱数据的光谱信息特点及最佳波段选择——以北京顺义区为例[J].干旱区地理,2001,23(03):214-220.

[5] 钱乐祥,等.遥感数字影像处理与地理特征提取[M].北京:科学出版社,2004.

[6] 刘建平,赵英时.高光谱遥感数据解译的最佳波段选择方法研究[J].中国科学院研究生院学报,1999,16(2):153-160.

26,其中小于通用土壤流失方程中规定的 LS 最大值 72 的地形因子值占总数的 95%。总体上看,结果是非常令人满意的,并且与实际情况吻合的较好。如图 3 所示,在山顶平台和水流汇集处,LS 值较低(白色区),相反,在等高线密集的陡坡之处(深色区),LS 较高。之所以出现较大的地形因子值,是因为所用的 DEM 数据精度低,不能详细表达地形的局部变化特征,加上水系数据也不够详细,致使累积坡长值过大,从而导致个别较高的地形因子值。

4 结论与认识

基于 DEM 通用土壤流失方程地形因子的计算方法可以充分利用 GIS 技术的空间分析功能,实现大范围区域地形因子的快速提取工作,最大坡降坡度的计算方法只考虑最大坡降方向的两个像元,保留了局部地形的变化特征,并且减少了计算量。累积坡长的定义较好的描述了自然界的实际情况,并且通过截止角度的引入和汇水区的叠加,标明了无侵蚀区域。同时,水系数据的引入大大提高了计算效率,使结果更加合理。DEM 数据精度是影响计算结果的重要因素,由于受计算机硬件设备的限制,通常 DEM 数据的精度不能太高,这就使得局部地形特征丢失,导致累积坡长的计算值偏高,从而影响最终的结果。需要指出的是,区域范围的大小与 DEM 数据精度之间的对应关系,即一定区域该用何种精度的 DEM 数据,仍需要进一步研究。

图像上,居民地信息为蓝色,能和背景地物较好的区分开,效果较好。实验表明,该方法操作简便,快捷,是进行假彩色合成的一种行之有效的办法。