

遥感科技资料(2)

# TM图象及其应用译文集

北京大学遥感技术应用研究所

一九八七年

## 致 读 者

遥感科技资料第二期——TM图象及其应用译文集，是在我所三位老师指导下，由85级研究生翻有关TM资料的习作，并由三位老师亲自校对，由于内容涉及学科较多，文章又较长，为此三位老师对每篇文章都作了部分摘译整理，但因时间短促，我们的业务水平和编辑力量所限，难免有译文不妥和错误之处，欢迎读者来函指正。

北京大学遥感技术应用研究所情报资料室

• 表示本期译文译自《IEEE Transactions on Geoscience and Remote Sensing》

1984年5月

第GE-22卷第三期

# 目 录

1. 与 MSS 数据相比 LandsatTM 数据优势的统计分析..... ( 1—11)
2. Landsat-4 专题制图仪和多光谱扫描仪数据的信息量对比 ..... (12—19)
3. 应用先进的图象处理技术分析和处理 Landsat-4 传感器数据..... (20—32)
4. Landsat-4 MSS 和 TM 数字图象数据特征..... (33—46)
5. TM 图象的特征 ..... (47—52)
6. 飞行中的专题制图仪绝对辐射率校准..... (53—58)
7. TM 影像质量, 配准、噪音和分辨率..... (59—68)
8. TM 数据缨帽变换的物理基础..... (69—76)
9. Landsat-4 专题制图仪几何特性分析 ..... (77—82)
10. 专题制图仪的辐射校正及几何纠正算法验证..... (83—91)
11. 利用航空专题制图仪资料识别海岸水体中的悬浮沉积物..... (92—99)
12. 估算内华达的弗吉尼亚地区航空 TM 图象中植被对 $1.65/2.22\mu$ 的  
    比值影响.....(100—107)
13. 用Landsat-4 专题制图仪推求雪的反射率 .....(108—112)
14. Landsat-4 专题制图仪和多光谱扫描仪资料对选定作物和  
    森林覆盖类型的光谱变异性.....(113—122)
15. 关于专题制图仪探测草原植被下土壤特性的评价.....(123—125)
16. 专题制图仪的地学应用研究介绍.....(126—164)

# 与 MSS 数据相比 Landsat TM 数据

## 优势的统计评价

DARREL L. WILLIAMS, JAMES R. IRONS, BRIAN L. MARKHAM,  
ROSS F. NELSON, DAVID TOLL, RICHARD S. LATTY, AND  
MARK L. STAUFFER

**提要**——为了定量地分析专题制图仪(TM)传感器对分类精度改进的效果,用1982年11月2日获得的哥伦比亚特区华盛顿幅的TM数据进行这方面的研究。这项研究工作是 根据 固定效应,三个因子,二级方差分析(ANOVA)而设计的。系统地降低TM数据在空间、光谱和辐射测量方面来模拟分别改变每一个传感器参数所产生的效应,并且结合其他传感器参数,最终模拟多光谱扫描仪(MSS)数据特征。

方差的最大值可用光谱波段变量来描述,它使分类精度平均增加 5.85%。该增量使 TM 数据相对于 MSS 数据提高了 20% 精度,即:相对改进百分率=(高精度值-低精度值)/低精度值 $\times 100$ 。方差的第二个最大值用“辐射”变量(即比特的量化水平)来表示。它使正确分类的象元百分比提高了 5.25%。由于量化水平的提高,它使 TM 数据相对于 MSS 数据提高了 19% 的精度。空间分辨率解释使分类精度普遍平均降低 0.7% 的基本原因,与 MSS 数据相比,它使 TM 数据相对降低了 2%。只有光谱波段组合提供的差异和量化水平在统计上是显著的, $\alpha$ 值在 0.01—0.001 之间。

## 引 言

Landsat-4 在轨道上带有两种遥感装置:多光谱扫描仪(MSS)和新型的传感器——专题制图仪(TM)。设计 TM 是为遥感领域提供一个改进的数据源。与 MSS 相比, TM 提供了较高的空间分辨率,新的和更适宜的波段,量化的辐射灵敏度达到八比特而不是六比特(表一)。传感器性能上的这些改进,希望能使数据的质量和信量大大地提高,从而增强这些数据在地球观测中的应用。因此在 Landsat-4 运行的第一年里,对 TM 传感器所获得的数据的质量和效应进行验证和定量分析是一个基本目的。

通常用来评价不同传感器装置所获得的数据的相对效应的方法是比较已达到的分类精度。在 Landsat 4 发射之前,使用安装在飞机上的专题制图仪模拟装置(TMS)所获得的数据进行了多次研究,定量分析 TM 数据与 MSS 数据相比在分类精度上的提高。然而,这些研究工作很少评价单个传感器的参数(如,光谱分辨率和辐射分辨率)在分类精度方面的效果或贡献。改变单个传感器性能的效应可以定性地测到。例如,增加光谱波段可以通过提供地物反射率在不同的光谱波段上差异把事先无法分离的地物种类区别出来。同样,改进

表一

Landsat—4TM和MSS 传感器的性能比较

| 波段名称   | 专题制图仪 (TM)          |                                    | 多光谱扫描仪系统 (MSS)      |                                    |
|--------|---------------------|------------------------------------|---------------------|------------------------------------|
|        | 波段( $\mu\text{m}$ ) | 辐射灵敏度( $\text{NE}\Delta\text{P}$ ) | 波段( $\mu\text{m}$ ) | 辐射灵敏度( $\text{NE}\Delta\text{P}$ ) |
| 1 波 段  | 0.45—0.52           | 0.8%                               | 0.5—0.6             | 0.57%                              |
| 2 波 段  | 0.52—0.60           | 0.5%                               | 0.6—0.7             | 0.57%                              |
| 3 波 段  | 0.63—0.69           | 0.5%                               | 0.7—0.8             | 0.65%                              |
| 4 波 段  | 0.76—0.90           | 0.5%                               | 0.8—1.1             | 0.70%                              |
| 5 波 段  | 1.55—1.75           | 1.0%                               |                     |                                    |
| 6 波 段  | 10.40—12.50         | 0.5K( $\text{NE}\Delta\text{T}$ )  |                     |                                    |
| 7 波 段  | 2.08—2.35           | 2.4%                               |                     |                                    |
| 地面瞬时视场 | 30米 (1—5,7波段)       |                                    | 80米 (1—4波段)         |                                    |
|        | 120米 (6波段)          |                                    |                     |                                    |
| 数据传递速率 | 85兆位/秒              |                                    | 16兆位/秒              |                                    |
| 量化等级   | 256                 |                                    | 64                  |                                    |

辐射分辨率和增加信噪比可以通过增加地物类型间界线而区分地类。相反,提高空间分辨率会对分类精度产生偏移影响。较高的分辨率可以降低混合象元的比例,因而提高分类精度。但是分辨率的提高使得地物类型内部的光谱差异增加,有碍于分类。

专题制图仪展示了一项颇为壮观的研究与发展的结果。其中,所有这些在传感器性能方面的主要改进都同时被纳入同一个系统中,反映了所有同步工作的传感器性能上的相互制约的关系,因此使得由每个单独传感器性能改进产生的影响的定量评价复杂化。然而,获得这种定量信息的能力对科学家和工程师确定传感器参数要求和为未来的遥感事业设计新的传感器将是非常有益的。

因此,位于马里兰州格林贝尔特的美国宇航局(NASA)哥达德飞行中心(GSFC)的地球资源部的一些训练有素的科学家,研究并定量分析每个TM传感器的参数对分类精度所产生效应的试验。本文讨论试验设计方案并概述使用1982年11月2日获得的华盛顿D、C.幅的TM数据所取得的成果。

## 试验场/数据说明

为了便于收集和野外验证地面参考数据,选定GSFC附近的一个地区作为一般试验区。它位于华盛顿D、C.以西,马里兰州Baltimore以北,Bheasapeake湾以东,马里兰州北Charles县以南。收集1982年7月13日在该地区上空拍摄的1:40,000比例尺的彩色红外航空立体象片,恰好在Landsat-4发射之前。

该地区有性质不同的城市、郊区和农业土地覆盖类型。其西部包含与华盛顿市郊毗连的许多住宅扩建区。整个试验场包括城市边缘的密集扩建区、郊区多户和单户的住宅区以及低密度的单户住宅扩建区。众多商店中心、工业联合企业,采石场和飞机场也分散于试验区

内。该区有Meade堡和安德鲁斯空军基地等军事目标。此外，该试验场还包括有农业用地和森林覆盖地区，农业用地基本上是分散的小田块，主要农作物是玉米，大豆和烟叶，还有牧场和草原。美国农业部 Beltsville 农业研究中心也位于该试验场内。就面积的大小而论，森林是一种主要覆盖类型，它基本由混交落叶林，针阔混交林和纯针叶林组成。该地区还包括与Chesapeake湾三角洲相连的低地植被群落。

整个试验地区七个波段的TM无云影响直到1982年11月2日才获得。TM和MSS影像是同时获取的，这两种数据集的质量都很好（如：无云覆盖，清晰度高、飞行器和传感器运行正常等等）。但由于11月份正是落叶林叶落色衰、大多数农作物枯萎、许多庄稼已经收割，此时获得的数据对于一般的分类判别并不十分理想。另外，由于太阳角度低，整个地区的总照度也降低。因此，TM和MSS传感器的整个动态范围和量化性能就不能加以利用。不过，已经决定着手进行这方面分析，因为：（1）它第一次提供了使用真实TM数据来进行设计和方法论的试验机会。（2）要求尽快定量评价与MSS相比TM传感器参数的特征。（3）可供选择的地区地面参考数据库不可能及时地、有效地建立起来。

## 步 骤

### A. 试验方案

这项工作的第一步是要设计一项试验，它将定量地评价增加的TM传感器的光谱分辨率、空间分辨率和辐射分辨率对分类过程所产生的影响，理想情况，要求具有单独地和一种可能的组合方式对每一个因子（即传感器的参数）所产生的影响进行评价的能力。Sigman和 Graig 的工作利用TMS数据，根据多因子的方差分析方法（ANOVA），提出一种有希望的结论。多因子方差分析方法适于TM真实数据的应用。因为该方法可以同时评价三个因子（即光谱、空间和辐射分辨率）在分类精度方面的效应。这里每一个因子都有二个值（即TM和MSS）。上述称为三因子固定效应的ANOVA方差分析方案。它需要对八个问题进行分析。

固定效应模型比随机效应模型更受欢迎。对于给定的自由度，固定效应模型为初级研究中的因子值（TM和MSS的因子值）提供了一个较灵敏的试验，固定效应模型的试验标准就是实验误差（即剩余方差）。随机模型既有实验误差还有其他干扰因素，因此测试灵敏度较差。然而，由固定效应模型得出的结论并没有提供因子值之间内插或者趋势概括。试验的单位是分类结果而不是象元。

试验方案所需要的八个处理方法，每一个都是由一组不同的、具有如下特征的数字影像一数据集组成：

- A 30米/8比特/6个波段（TM真实数据）
- B 80米/8比特/6个波段.
- C 30米/6比特/6个波段.
- D 80米/6比特/6个波段.
- E 30米/8比特/3个波段.
- F 80米/8比特/3个波段.
- G 30米/6比特/3个波段.

#### H 80米/6比特/3个波段 (由TM模拟的MSS数据)

(注:下面划线的条目是那些相对于处理方法A有变化的因子, 处理方法A由TM的真实数据组成。)

八个处理方法中的每组数据集都是通过压缩TM数据(处理方法A)产生, 以便接近MSS数据每个因子的特征(例如, 降低TM的空间分辨率, 大致接近MSS的空间分辨率)。然后, 对这些数据集统一进行分类。多因子ANOVA方案可以对处理方法间精度的显著差值进行统计试验, 从而就可以对单个传感器的改进和其他因子的组合改进在分类实施中产生的效果进行定量评价。

#### B. 资料准备与分析

资料准备和资料分析包括: 航空象片解译及结果数字化, 以获得数字地面化的参考数据集; 试验场的TM数据的提取, 为模拟ANOVA方案所需要的MSS波段的空间分辨率和辐射特征而进行的TM数据的预处理; 每种覆盖类型的训练区和试验象元的选择; 数据的聚类与分类, 对每组数据集得出的分类精度评价。

#### 象片解译与数字化:

1:40,000的彩色红外航空象片是于1982年7月13日拍摄的, 恰好在Landsat-4发射时间之前。为了进行细致的分析, 随机选择了九幅象片。在象片解译之前, 为了把航空象片边缘的典型畸变产生的影响降到最低限度, 从241mm(9英寸)规格的象片的象主点起, 取其上下左右各50mm长作为试验场子区的象片, 这块边长各为100mm的方形区域大致占整幅象片20%的中心地区, 约合实地面积1600公顷。然后将1:40,000比例尺的子区象片照像放大四倍, 成为常规比例尺1:10,000的象片, 并把它印刷成图。在每张象片上蒙上一张醋酸酯透明薄膜, 这样可以把解译结果直接画在薄膜上。为了区分各种土地覆盖类型, 所用最小的制图单位为15米(即TM瞬时现场的一半)。然而仅就农业用地而言, 所用的最小制图单位仅仅是为了把田地相互区别开来, 而不必描述田块内部的差异。用这种方法就可以判读出17种第二和第三级的土地覆盖类型和土地利用类型(表二), 画出地类界, 标出注记。必须注意的是土地覆盖或土地利用类型并非总是互不相容。因为对于空间尺度变化较大的区域(如居民地)就无法严格地按照15米的最小制图单位。例如, 在下面这种情况使用土地利用类型而不是土地覆盖类型, 土地覆盖的种类组成(如: 房顶, 草坪, 树, 居民地周围的混凝土或沥青地块)占据空间尺度小于或近似等于15米最小制图单位的区域。在10月份的最后一个星期内, 即在TM数据获取之前, 通过野外调查和计算, 对象片解译和注记结果进行了核实与修正。

对在醋酸酯薄膜上画出的多边形进行数字化, 并在几何上配准到TM数据。把TM数字影像用做制图参考, 以避免影响TM数据重新采样。配准通过匹配TM数据上和象片上的控制点来完成。控制点标在醋酸酯薄膜上, 然后输入数字化程序。整个数字化过程和配准过程都是通过一个具有综合的、自动数字化的软件和硬件设备的小型机器(芝加哥航空调查局生产的)来完成。就是把每一张薄膜放在分辨率为每毫米40线的滚筒式扫描器上, 该扫描器就把多边形用格网化数字表示出来, 每一条数字化线都被压缩成一个格网元的宽度。把格网化数据转换为矢量数据, 然后利用地面控制点配准到TM数字影像数据。这种方法避免用手扶式光标和数字化台多边形人工数字化过程中的固有误差。

产生矢量数据后, 在对话式编辑工作台把每一条数字化的多边形都用适当的土地覆盖或

表二 在地面参考数据中划分的17种土地覆盖和土地利用类型一览表

| 土地覆盖类型              | 闭度为50—100%                      |
|---------------------|---------------------------------|
| 1. 水                | 11. 硬木林：占森林成份70%以上林冠郁闭度为10—50%  |
| 2. 混杂农作物            | 12. 针叶林：占森林成份70%以上，林冠郁闭度为10—50% |
| 3. 玉米               | 13. 混交林：林冠郁闭度为10—100%           |
| 4. 玉米茬              | 14. 沥青地                         |
| 5. 灌木丛              | 土地利用类型                          |
| 6. 草场/牧场            | 15. 单户居民住宅区                     |
| 7. 大豆               | 16. 多户居民住宅区                     |
| 8. 裸地—耕地            | 17. 工业区/商业区                     |
| 9. 裸地—空地            |                                 |
| 10. 硬木林：占林冠70%以上林冠郁 |                                 |

土地利用类型标注。这些专题数据集随后变换到每一幅经解译的象片的二个格网化数据文件里，其中一个格网化文件里存放代表28.5米×28.5米面积象元。然后这个文件可以用数字形式重迭到具有TM空间分辨率特征（即处理方法A，C，E和G）的四组数据集上。另一个文件中存贮了象元面积为57米×57米的数据，它可以直接重迭到空间分辨率被降低的数据集（即处理方法B，D，F和H）上。因此，这些以格网形式配准的、数字化的专题数据集就被用作地面参考资料。

TM数据的提取与预处理：与九个随机选择的试验场相对应的TM数据，从该地区原始的P格式数据磁带中提取出来，以简化下道程序，所提取的每一个TM数据块，大小为256×256个象元，包括分辨率为30米的六个波段的全部数据。地面分辨率为120米的热红外数据，由于其空间分辨率的显著差异，故在该项研究中不予考虑。

为了从每一个传感器参数上来模拟MSS的特征，ANOVA方案要求在光谱上、空间上和辐射上降低TM数据要求。然而，由于TM和MSS在光谱段划分上、辐射响应、横向跟踪扫描方式等等方面存在着的内在差别，由TM数据不可能精确地模拟MSS的辐射、光谱波段和空间分辨率。因此，在降低TM数据来模拟MSS数据特征的过程中，需要进行一定的折衷或简化假设，其步骤如下：

(a) 光谱分辨率模拟：MSS的光谱模拟是简单利用TM的2、3、4波段实现的。这些波段对模拟MSS的所有波段是最有效的组合波段。TM的第二个波段（0.52—0.60 $\mu$ m）非常接近MSS的第一个波段（0.50—0.60 $\mu$ m）。TM的第三个波段（0.63—0.69 $\mu$ m）与MSS的第二个波段（0.6—0.70 $\mu$ m）非常接近。TM的第四个波段（0.76—0.90 $\mu$ m）包括了MSS的第三个波段（0.70—0.80 $\mu$ m）和第四个波段（0.80—1.1 $\mu$ m）的一部分波段区域。由于这两个波段都覆盖与生物量有关的0.74—1.10 $\mu$ m光谱区主要部分，对植被研究目的，TM第四波段被认为是最接近MSS第四波段的。MSS第三波段包括了由叶绿素吸收引起的植被低反射率（约在0.71 $\mu$ m处终止）与植被高反射率（约在0.75 $\mu$ m处开始）之间过渡光谱区。由于此

因, MSS第三波段对于植被识别就不是最佳的波段。

(b) 辐射分辨率模拟: 辐射分析就是对量化水平进行比较。进行MSS的辐射模拟, 是把TM数据的0—255个电位箱 (potential bins) 变换到MSS数据0—63个电位箱 (即把TM的每个数据四等分, 四舍五入为最接近的整数值)。采用这种方法, 不必说明两种传感器之间的动态范围和信噪比的差异, 它只简单地说明六比特量化水平与八比特量化水平的问题, 而不是TM辐射灵敏度与MSS辐射灵敏度问题。

(c) 空间分辨率模拟: 模拟MSS的空间分辨率, 通过计算TM数据  $3 \times 3$  象元窗口内的简单算术平均值进行。窗口每次移过二列以模拟80米分辨率的MSS57米沿扫描方向的采样速率。在移过一幅影像后, 均匀的窗口增加两行以近似MSS数据的几何重新采样, 把沿飞行方向80米的正常象元尺度压缩到57米。这样, 就模拟出由 EROS 所确定的标准MSS P—磁带上57米 $\times$ 57米象元规格。

由于Landsat-4MSS数据和TM数据是同时获得的, 就可以评价上述实际模拟MSS传感器过程。通过比较H数据集得到结果 (即从TM模拟MSS的数据) 和分析MSS的真实数据得到结果, 就能评价用压缩的TM数据模拟MSS特性过程的有效性。例如, 如果模拟工作是完善的, 可以希望用各数据集得到的分类精度是一致的。因此, 为了评价上述模拟过程合适程度, 可在ANOVA方案的八个数据块中增加一组数据集——第九组数据集, 叫处理方法 I, 由MSS的真实数据组成。处理方法H (模拟的MSS数据) 与处理方法 I (MSS真实数据) 比较的结果, 将在“结论与讨论”部分论述。

**训练区与试验象元的选择:** 对于17种土地覆盖和土地利用类型, 每一类型都需要从ANOVA试验方案中的八组数据集中选出训练区和试验象元。从给出的数据集中选出样本, 用来聚类、分类, 及对那组数据集分类的精度结果造表。为了简化统计试验及同等对待每种类型, 不考虑某种类型在该试验区出现的频率, 对所有的土地覆盖和土地利用类型均需要相同数量的训练区和试验象元。

利用前面提到的数字化地面参考数据自动选择样本象元。编写软件以在地面参考数据中从每种土地覆盖和土地利用类型随机地选出相同数量的象元。对57米分辨率的数据, 每种类型选择550个象元。对28.5米分辨率的数据, 选2200个象元。样本的大小受最小地面参考类型的实际象元数的限制 (如: 大豆是最小的覆盖类型, 在57米分辨率的数据中, 它含有571个象元。在28.5米分辨率数据中, 它含有2244个象元)。

对ANOVA方案中的每一个处理方法, 用这种方法产生两组随机样本。对每一组样本, 通过对样本的不同项目进行训练和试验, 得到两组相同的样本。因此, 对于一个给定的处理方法, 两组样本共有四个相同的样本。对于一个给定的复制样本, 在每种类型随机选择的550个57类分辨率的象元中, 取出400个用来聚类以得到训练统计值。把剩余的150个象元根据训练统计值对所有类型进行分类, 这些结果用来试验分类精度。训练象元 (2200个中的1600个) 与试验象元 (2200个中的600个) 比值用于28.5米分辨率的数据集。

**样本数据的聚类与分类:** 上面讲到的样本数据的训练统计值是用ISOCLS聚类算法得到的。聚类是分别对来自每个处理方法的每一组训练象元进行的。建立ISOCLS程序参数, 以便使该程序实现迭代的点移动的聚类算法的功能, 其中, 最初聚类的均值随着每一次迭代移向自然光谱集群的矩心。最后得出的训练统计值没有进行整理 (即没有对光谱类别进行合并或删除), 目的是为了防止把分析者的偏见代入分类结果中。虽然没有对训练统计量进行

整理限制了分类的精度，但是聚类过程的一致性可以更有效地比较处理方法之间的分类精度。

总之，进行544个不同的聚类过程（八个ANOVA处理方法，每一个处理方法四个复制样本，每个复制样本包含17种覆盖类型产生了32组不同的训练统计量。然后对试验象元的每个样本用适当的训练统计值组和逐点高斯最大似然分类器进行分类。从试验象元获得分类精度值。

为正确表示出正确分类试验象元的数目，建立列联矩阵同时产生删除误差和组合误差。每个矩阵对角线上的值（即每种覆盖类型中正确分类的试验象元数）的总和被试验象元总数去除，就得到总的分类精度值。这个响应变量——即正确分类的象元百分比，就是精确的呈双正态分布的随机变量。对计算分类精度值作反正弦变换，使处理方法之间的方差相等，并把这些数据转换成一个正态分布。在每一个处理方法的数据中，因为分类精度值要重复使用四次，所以每个处理方法中，四次变换的精度值用来进行三个因子的ANOVA。采用SAS ANOVA步骤进行这种分析，它可以分析单因子的效应和因子相互作用效应。

## 结 论 及 讨 论

表三中总结ANOVA方案中对八个处理方法所得到的全部分类的精度值。对于每一个处理方法，表三列出了在“聚类与分类”部分中所讨论的四个复制样本象元进行正确分类的平均精度值和变换精度值的标准方差。全部精度值从37.9%（处理方法B：80米/8比特/6个波段。）到25.7%（处理方法G：30米/6比特/3个波段）。最大精度范围是12.2%，所有处理方法的平均精度为31%。

这些精度非常低的数字是有意义的，全面、仔细观察试验方案和数据过程，可看出在选择的数据分析中没有不合适的方面。记住，试验的目的是对因数据特征带来的分类精度方面的差异进行准确的统计评价，并非增大分类的精度。低精度值是由下列因素造成的：1)试验方案包括缩小分析工作者的偏差，使分类精度的差异完全是由于数据特征的差异造成的，而不是数据分析方法上的偏差造成的。2)勾画17种被详细描述的第二级和第三级土地覆盖和土地利用类型，大多数代表植被覆盖的条件，它们在光谱上相互非常接近，这是由于数据获取是在同年同一时间。3)训练区和试验样本象元的选择完全受数字地面参考数据的控制，因此“边界”的象元和“田块中”的象元都被选上，而且任何在参考数据向TM数据中产生的微小误差，都可能导致混杂的样本数据的产生。

为了检验分析方法与结果的有效性，要粗略地研究一下下列试验的效果。（1）允许分析者在分类之前用对话方式编辑训练统计量值。（2）单一地减少土地覆盖/土地利用类型的数目。在第一个试验中，采用标准方法对处理方法A（即TM真实数据）重复分析，这里允许分析者利用计算机人机对话来编辑训练统计值，不选择新的训练场，仍保留17种土地覆盖和土地利用类型。。分类精度值就从36.7%上升到62%（表三）。在第二个试验中，把17种第Ⅱ/Ⅲ级的类型合并成五种第Ⅰ/Ⅳ级的类型（即：水、农作物、牧场和草场、森林及城市），并计算出每种处理方法的分类精度值以及所有问题平均精度值。训练或分类中不允许分析者进行人机对话操作。对划分五种类型的精度值为59.5%到71.7%，其平均值为65.7%。它们可以分别与17种类型的相应精度值对比：25.7%，37.9%，31%（表三）。这两项增加

表三、作为传感器或数据特征函数的整体分类精度及被划分的土地覆盖类型

| (PPC—正确的分类百分比)        |            |                     |                       |                            |      |
|-----------------------|------------|---------------------|-----------------------|----------------------------|------|
| 问题 特 性                | 情 况        |                     |                       |                            | 情况 I |
|                       | ppc平<br>均值 | 反正弦变<br>换ppc平<br>均值 | 反正弦变<br>换ppc的<br>标准方差 | 相对处理方法<br>A有变化的<br>ppc的平均值 |      |
| A 30米/8比特/6个波段        | 36.7(62)*  | 37.3                | 0.43                  | N.A.                       | 71.7 |
| B 80米/8比特/6个波段        | 37.9       | 38.0                | 0.67                  | + 1.2                      | 71.5 |
| C 30米/6比特/6个波段        | 31.2       | 34.0                | 0.79                  | - 6.8                      | 68.4 |
| D 80米/6比特/6个波段        | 29.9       | 33.2                | 0.66                  | - 6.8                      | 63.0 |
| E 30米/8比特/3个波段        | 29.0       | 32.6                | 0.24                  | - 7.7                      | 65.1 |
| F 80米/8比特/3个波段        | 30.9       | 33.7                | 0.94                  | - 5.8                      | 64.4 |
| G 30米/6比特/3个波段        | 25.7       | 30.5                | 0.38                  | -11.0                      | 59.5 |
| H 80米/6比特/3个波段        | 26.7       | 31.1                | 0.63                  | -10.0                      | 62.2 |
| I Mss真实数据-1.2.4<br>波段 | 20.5       | N.A                 | N.A                   | -16.2                      | 54.9 |

(注：下面划线的条目是相对A即TM真实数据有变化的那些因子)

• 允许分析者进行人机对话操作时编辑训练统计量值得到的精度值。

1 根据17种分类了的土地覆盖和土地利用类型的用途所得到的精度值。

2 根据5种合并了的土地覆盖类型的用途所得到的精度值。

的试验得到的分类精度值对给定试验方案的地面规律和收集数据的时间无疑是比较好的。

如前所述，对降低TM数据以模拟MSS特征技术的有效性和合适程度进行评价。在这种情形下，把1982年11月2日过境时所获得的真实MSS数据配准到模拟的数据集中去，并且进行与MSS模拟数据同样的分析过程。与MSS模拟数据的26.7%分类精度值相比（表三），MSS真实数据的分类精度值为20.5%。二者精度相差6.2%正是所期望的，并认为它是下列事实的结果：1) 与模拟数据相比，真实MSS数据中加了一个配准和重复采样的过程，导致内插象元值和配准误差。2) 在压缩过程中有个已知的限度，如不可能完全消除信噪比。因此，TM的压缩数据在信噪比方面仍比MSS的好些。3) 空间分辨率的压缩采用平方阶或门函数，而MSS真实点扩散函数更接近高斯函数。4) Landsat-4的MSS数据质量受系统中的相干噪音影响多少有些降低，因此与TM的压缩数据相比，它在可达到的分类精度方面就有影响。因此由模拟提供的方差估计可以认为是保守的。直接比较TM的真实数据和MSS的真实数据（即表三中的处理方法A与处理方法I）表明，与由模拟MSS数据得到10%的改进相比，用TM数据达到了16.2%的改进。

在分类精度质造表分析之后，对处理方法精度问题进行方差显著性统计试验。其结果列于表四，表五，表六和表七中。重申一下，固定效应ANOVA方案中评价的因子是：空间分辨率（30米与80米），光谱波段数（六与三）和量化水平（八比特与六比特）。每个因子有两个值，反映了TM传感器和MSS传感器的特征。

表四列出了方差分析的结果。显著性效果是由F试验决定的（处理方法I误差的概率为

0.05,  $\alpha=0.05$ )。三个因子交叉项(表四)并不显著,而二个因子交叉项的效果却都显著,尤其是光谱与辐射交叉项。在从六比特量化水平的三个波段增加到八比特量化水平的六个波段时(即30米分辨率时为7.7%与7.1%;80米分辨率时为5.5%与3.3%。见表五),这个光谱与辐射的交叉项基本表明分类精度大大提高。

表四 正确分类象元百分比的反正弦变换结果的方差分析

| 因 子      | DF | 平方和     | F 值         | $\alpha$ 有效位数 |
|----------|----|---------|-------------|---------------|
| 空 间      | 1  | 1.347   | 3.53        | 0.0716        |
| 光 谱      | 1  | 105.435 | 266.23      | <0.0001       |
| 辐 射      | 1  | 83.555  | 210.99      | <0.0001       |
| 空间与光谱    | 1  | 1.687   | 4.26        | 0.0500        |
| 空间与辐射    | 1  | 2.106   | 5.32        | 0.0300        |
| 光谱与辐射    | 1  | 6.004   | 15.16       | 0.0007        |
| 空间与光谱与辐射 | 1  | 0.512   | 1.29        | 0.2667        |
| 总 模 式    | 7  | 200.698 | 72.40       | <0.001        |
| 误 差      | 24 | 9.505   | MSE = 0.396 |               |

表五 用Tukey法成对比较反正弦变换中光谱因子产生的正确分类的处理方法平均值百分比(Pcc)的差异

| 处理对 | Pcc差异 | 反正弦变换中Pcc的差异 | $\alpha$ |
|-----|-------|--------------|----------|
| A—E | +7.7% | +4.7%        | <0.01    |
| B—F | +5.5% | +4.3%        | <0.01    |
| C—G | +7.1% | +3.5%        | <0.01    |
| D—H | +3.3% | +2.1%        | <0.10    |

表六 用Tukey法成对比较反正弦变换中辐射因子产生的正确分类的处理方法平均值百分比(Pcc)的差异

| 处理对 | Pcc差异 | 反正弦变换中Pcc的差异 | $\alpha$ |
|-----|-------|--------------|----------|
| A—C | +5.5% | +3.3%        | <0.01    |
| B—D | +8.0% | +4.9%        | <0.01    |
| E—G | +3.3% | +2.1%        | <0.10    |
| F—H | +4.2% | +2.6%        | <0.01    |

表七 用Tukey法成对比较反正弦变换中空间因子产生的正确分类的处理方法平均值百分比 (Pcc) 的差异

| 处 理 对 | Pcc的差异 | 反正弦变换中Pcc的差异 | $\alpha$ |
|-------|--------|--------------|----------|
| A—B   | -1.2%  | -0.7%        | >0.10    |
| C—D   | +1.3%  | +0.8%        | >0.10    |
| E—F   | -1.8%  | -1.1%        | >0.10    |
| G—H   | -1.0%  | -0.6%        | >0.10    |

注:  $\alpha$  是类型 I 的误差概率

由于因子交叉项的显著性使得模式中的主要效果的解译不太清楚,所以就成对地比较单一处理方法的平均值。与只有一个因子情况不同,处理方法之间分类精度的差异用多次比较的Tukey方法来分析,根据其它因子把光谱段从三个增加到六个,使得分类精度从3%增加到8%,而且所有的差异都显著(表五)。把量化水平从六比特增加到八比特,也使分类精度从3%提高到8%。所有的差异显著地大于0(表六)。然而,把空间分辨率由80米提高到30米,使分类精度的混合效应在 $\pm 2\%$ 之间,以致精度的差异不显著。

这些结果指出,改进量化水平和增加可见光及中红外谱区的光谱波段将导致用逐点高斯最大似然分类器对土地覆盖类型精确分类能力的提高,另一方面表明,空间分辨率提高到30米,并不能明显地提高分类的精度。这个结论也许出人意外,因为把同时获得的TM和MSS图象在一起目视比较。可以看出,TM数据包含有更多的空间细节。事实上增强效果很好的TM子图象很容易与高空摄取的航空象片混淆,改进空间分辨率产生的“不显著性”(insignificance)可能与这一事实有关,即试验是围绕着逐点最大似然分类器设计的。该分类器仅仅是根据象元的光谱特征进行分类的,而不全盘考虑纹理、形状、相关布局、组合关系或与相邻象元有关的空间特征信息。另一个可能影响空间分辨率显著性的因素是与地面参考类型有关的详细程度。例如在MSS数据中地物内部的差异并不明显而在TM数据中却很明显,而且地面参考数据没有区分这些有差异的类型。这些差异常常代表了目标物实际物理特性。把数据合并成大的种类,不可能充分利用TM数据的信息优势。这些因素不可能使结果倾向于有利于MSS数据,而不完全体现TM数据空间分辨率的优越性,这在TM与MSS图象目视比较中已经明了的。鉴于这些结果,我们研究的重点应该放在对高空间分辨率的数据特征的全面认识方面,以及更好地利用这些特征来用于信息提取的优势算法。遥感领域对遥感数据的处理、分析和应用的认识原先是围绕MSS数据的性能和特征设计的,现在需要改变。这可能包括对地类和地类特征的重新定义,就象大的覆盖类型不适用于高分辨率的数据。如果对高分辨率的数据没有进一步的理解,那么数据分析能力就比来自具有高分辨率的新传感器更有利于来自8比特量化水平,增加的划分好的光谱波段的新传感器的数据。

## 总 结

为了单独或一起定量地评价每个传感器的参数在信息提取过程中的影响,进行了固定效应、三个因子(每个因子有二个值)的方差分析的试验。被检测的传感器的三个参数是:

1)光谱波段数; 2)辐射量化值; 3)空间分辨率。试验所用的 TM 数据是1982年11月2日在华盛顿 D.C. 获取的。为了模拟MSS数据,把原始的TM数据在光谱、空间和辐射特征上进一步压缩。分类精度是以 ANOVA 方案中全部八种处理方法用一致的方法获得的。对处理方法间变换的分类精度差异显著性统计试验表明,由 TM 传感器设计提供的光谱波段的增加和量化性能的改进,将导致可以达到的分类精度相对于MSS的显著改进。然而,与其相反,由 TM 传感器提供的空间分辨率提高并不能提高分类的精度。这后一结论被认为是目前所用的分类算法(即逐点判断标准)的函数,而不是高空间分辨率数据带来的益处(或它的缺陷)的最后说明。这一点可通过对同时获得的 TM 和 MSS 图象的简单解译得到证实。因此提高对高分辨率数据的认识和分析,应先于或同时研制具有与 TM 相同或高于它的空间分辨率的新的遥感装置。如果开展这方面的工作,那么研究的重点就应该放在光谱波段的数目和要求及提高辐射测量灵敏度及其量化水平。

焦 健 译•

王敏华 校

# Landsat-4 专题制图仪和多光谱扫描仪 数据的信息量对比

JOHN C. PRICE

**摘要** Landsat-4专题制图仪(TM)和多光谱扫描仪(MSS)获得的同步数据使得可以对这两种图象数据在技术性能和数据应用方面进行对比。本文用五幅农业区图象对可见光和近红外波段的数据所含“信息”进行评价,得出TM在信息收集能力(用每象元比特数或单位面积比特数表示)上有显著优越性。TM的六个反射通道每象元从可能的 $6 \times (8 \text{ 比特}) = 48 \text{ 比特}$ 中获得18比特信息量;四个MSS通道每象元从可能的 $4 \times (7 \text{ 比特}) = 28 \text{ 比特}$ 中获得10比特信息量。因此TM和MSS在获得信息上具有同等的效果( $18/48 = 10/28$ ),这样的结果归因于:(1)TM光谱通道选择优越;(2)TM数据的高精度,即较低的系统噪声;(3)高空间分辨率的优点,即使在农业区,一个田块尺寸大于一个MSS象元尺寸。

因为MSS缺少热红外通道,所以对TM空间分辨率为120米的10—12微米波段的数据用能量平衡法进行理论分析。TM热红外数据主要对白天温度不变的水体制图、估算表面湿度和监测与人类活动有关的热特征等方面有价值。

## I 引 言

本文研究是用五幅农业区同时获得Landsat-4的TM和MSS数据进行统计对比。

以前很少比较不同卫星数据类型是有以下两个主要原因:1)在以前的卫星中,空间分辨率和光谱通道的数量、位置不同等;2)不可能获得代表同一观察镜下准确的同时同地数据集。以下对TM30米分辨率的6个光谱通道和MSS80米分辨率的4个光谱通道获得的信息进行比较。

在卫星观测系统中,熵与“0”和“1”之间的能量差别有关,这里很重要的是保证电子串扰、宇宙射线效应、热波动等因素,在传送到地面、计算机处理等过程中,不以随机方式改变Landsat数字值,这些数值记录在观察仪的探测器上。因此熵与数据传送和处理的可靠性和再现性有关,而不是与数据本身的数字值有关。所以如果观测系统设计得好,在经处理的数据中就看不出熵值变化带来的影响。

本文中信息引用由Shannon提出的有关这个课题的经典论文中的概念。“信息”对应于统计可变性,被定义为

$$H = - \sum_i P_i \log_2 P_i \quad (\text{比特}) \quad (1)$$

式中 $P_i$ 是在一系列测量中某一给定数字值 $i$ 的概率。这个定义既适用于连续数据也适用于离散的(整型的)数据,但普遍适用于通讯系统开发中的数字数据,图2—1以8比特表示的数

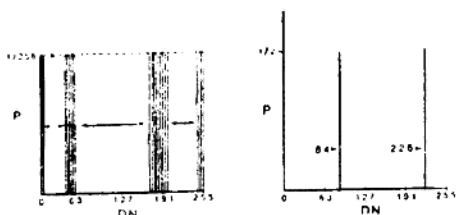


图2-1a: 8比特数据直方图(在0—255范围内所有数值发生的概率相等)  
b: 1比特数据直方图(只标出256种可能性中的两种)

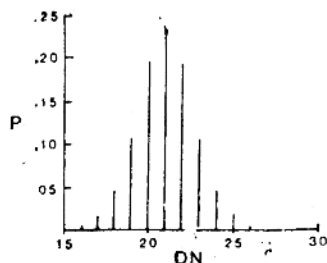


图2-2 数据值携带2.75比特数据直方图

据集直方图和一个携带1比特信息的数据直方图, 两图中0—255横坐标表示有携带8比特信息的数据潜力, 1比特数据原理上是能够在编码 $DN_{84}=0$ ,  $DN_{226}=1$ 后储存在计算机或磁带中或以每个数据值1比特进行点到点的传输, 这样的数据压缩技术减少了数据在单元间传送中所需的带宽, 并且减少储存在磁带中的数据的体积等。图2—2说明每次量测携带2.75比特信息的数据直方图, 类似 Huffman 创造的程序可以用来压缩隐含在信息量计算中的数据。

当考虑多个数据通道时, 如多光谱图象数据情况, (1) 式中的信息的定义可以扩展。信息符合通道间的可变性。例如, 如果已知两个光谱通道的测量值相等, 那么第二个值没有携带信息, 可以从数据集中消除。

这种多余信息本文是通过主成分变换来处理的, 即对变量的线性变换得到原数据集的一个更有效和简单的表示法。如果时间上或空间上的相互联系存在于数据流中, 例如沿一个图象行从一列到一列采样, 那么信息量的进一步减少是不言而喻的。图2—3图理想化的形式说明这种相互关系, 曲线近似于加利福尼亚因皮里尔河谷的一段扫描线, 这地区田块面积很大。图2-3还说明扫描宽度的代码过程, 其中数据后面跟着重复因子。

专题制图仪带有另一个热红外通道 ( $10.4-12.5\mu m$ )。在 MSS 上没有相应的通道, 这个通道的数据空间分辨率是120米而 TM 其它6个通道分辨率是30米。另外这个光谱段反应不同的物理效应。

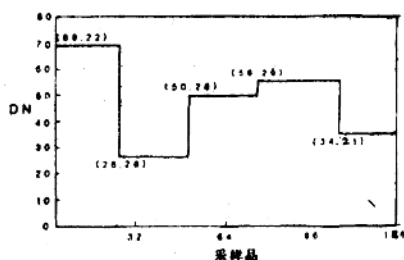


图2—3 沿一个有大量样本间多余信息的假设图象线的数据值, 每个值都有重复因子。

## II TM 和 MSS 反射数据的信息量对比

根据 Landsat—4 运行早期几个月中获得 5 幅相配的图象进行分析。表一列出了各个图象对的地点、日期和图象号, 这些图象是在美国作物生长晚期获得的, 代表美国农业区合理的地理取样, 从表一中图象表示的变异性要大于仲夏时出现的变异性, 那时农田区变成了绿色植被的海洋。

象许多地球物理数据一样, 陆地卫星数据的变异性 and 由此产生的信息量同时依赖于给定地区的地理状况及地区的面积大小。如果地区的大小保持不变, 就会发现在不均匀地区如一

表一 用于Landsat—4专题制图仪和多光谱扫描仪对比的图象

| 地 点         | 图 象 号       | 获得时间       |
|-------------|-------------|------------|
| 阿肯色东北部      | 40037—16031 | 1982.8.22  |
| 衣阿华西北部      | 40072—16325 | 1982.9.26  |
| 肯塔基西部       | 40094—15574 | 1982.10.18 |
| 华盛顿特区       | 40109—15140 | 1982.11.2  |
| 加利福尼亚因皮里尔河谷 | 40149—17444 | 1982.12.12 |

个城市即在中国或印度境内划分很细的农业区的单位面积信息量要比在不均匀地区如沙漠或草原的单位面积信息量大。因此对比MSS和TM重要的是要选择相匹配的地区。

另一方面,很明显随着地区的增大,信息量(比特数/面积)也将增加。因此如果开始用一个大小适度的农田区,并且稳步扩大所评价地区的面积,那么这个地区最终将包括山脉、沙漠、森林覆盖的赤道地区、海洋、极地冰盖等。这种不断增加的地理差异将使数据变化相应增加,并导致信息量的增加,因此在对比Landsat—4TM和MSS的能力时,必须选择一个固定大的或小的地区。

在本文的分析中,选择相配地区代表MSS $256 \times 256$ 象元,代表TM $512 \times 512$ 象元。因为MSS数据的空间分辨率是80米,经处理达到57米,所以最后数据的分辨率与处理过的TM数据的28.5米分辨率的比不完全等于2:1,原则上可以用MSS的“A”数据(经过校准,但没有重新采样)代替“P”数据(几何纠正中经过校准和重新采样)与TM进行对比,但是TM和MSS其他差异如精度(8比特对6比特)等,也使整个问题复杂化。本文中每个子区相当于边长为14.6公里面积为21300公顷的正方形。

由于历史原因,NASA对TM光谱波段序号没有按顺序命名。因此NASA的排列序号(1, 2, ..., 6, 7)所对应的波长值可粗略地表示为0.4, 0.5, 0.8, 1.5, 11, 和2.2微米,即6和7波段没按顺序排列,本文采用符合逻辑顺序的波段(L),所以NASA的6和7波段分别变成L7和L6波段,在本节的分析中略去L7热红外波段。

表二是表示5幅图象中每幅的三个子区的平均值信息量的对比。这些数值是对用户磁带上提供的已校准数据而言的。在每幅图象的三个子区中,某些地区间的偏差是显著的。

这些值可以与TM每通道可能的8比特信息量和MSS每通道可能的7比特信息量比较。MSS数据以6比特数据来传送,然后“扩充”,即把0—63的数值范围通过非线性查找表变换到0—127范围。MSS数据直方图的研究表明,探测元件间的偏差产生了引起查找表偏移的校正误差,结果充填了绝大部分可能由简单1—1变换带来的直方图中的空穴。所以数据出现7比特精度。

所有的H值都比理论极限值要小这一事实正是所其望的。因为对两个系统来说饱和辐射校准到典型地面反射水平以上,即使是在高的太阳入射角情况下,这里讨论的影响是在夏季和秋季得到的,而最大日射纬度是在北纬20—40度,MSS和TM的数据值普遍降低到可能范围的一半以下,这表明理论上能达到的信息量值相应减少。亮云能使H值大大增加。当然这样的数据对于土地应用是没有用的,在选择子区时要避开景象中出现的少数云区。