

陆地卫星数据信息提取的一个新方法*

曾志远

(中国科学院土壤研究所)

1988年6月14日收稿

摘 要

陆地卫星图像的磁带数码值或扫描密度值,均可被转换成绝对辐射值。四个波段中的每一波段的辐射值,均可分解为具有不同物理意义的三部分,其中,四个波段中物理意义相同的那一部分,又可被重新组合起来,于是,便得到三个新变量。它们分别被命名为总辐射水平值 L , 可见-红外(辐射)平衡值 B , 和波段(辐射)变化向量值 V 。 L 能监测地物的辐射强度; B 能监测水、水分和地表的潮湿程度,并能区分云、雪和冰; V 能监测植被的密度与生活力,因而是植被状况的良好指标。这三个变量几乎包含了原四个波段的全部信息,因此,通常只需处理这三个变量。算法基于回归分析且维数低,因此十分简单。因而,可节省图像处理时间和费用。使用这个方法,可以获得真正三色的、接近于自然色彩的真彩色合成图像,还可获得相应于地物特征的有意义的密度分割图像。此方法已被应用于中国和世界范围内的遥感资料。

关键词 图像处理 信息提取 特征提取 LBV 变换 KL 变换 KT 变换 磁带数码值 密度值 辐射值 辐射定标 辐射值换算 彩色合成 密度分割 回归分析

一、引 论

十余年来,已经提出了不少关于陆地卫星数据信息提取的方法^[1-10]。其中最重要的为 Karhunen-Loève 变换(KL变换)和 Kauth-Thomas 变换(可简称为KT变换)。前者以主成分分析为基础,后者以 Kramer-Schmit 变换为基础。二者均有较广泛的应用。

但是,KL变换计算复杂,需要进行大量的矩阵、逆矩阵、特征值、特征向量等复杂运算。而且其变换方程的系数在每一次图像处理时都需重新确定。KT变换计算虽较简单,但其设计主要是针对特定的目的,即农业生物量的监测,而且其计算方程的系数又依赖于特定地点两个原始点的选择^[11],这样又限制了它的应用范围。

此外,KL变换、KT变换以及以往的其它信息提取方法,基本上都是先用已有的数学统计方法(主成分分析等)对图像数据进行运算,而不管其结果如何,只有在通过运算得到新变量之后,将新变量与实际地物对照,才能大致确定这些新变量的意义。可以说,这些方法在很大程度上是纯数学式的,带有很大的盲目性。

本研究提出的新方法,是先分析实际地物的遥感特性(辐射特性),然后再考虑已有的数学方法,看是否有一种数学方法能提取出地物的这些遥感特性。

* 此研究为作者 1985—1987 年在加拿大遥感中心 (CCRS) 和加拿大土地资源研究中心 (LRRRC) 完成的研究工作的一部分。

这样设计出来的信息提取方法,运算提取的目标明确,得到的新变量物理含义清楚,计算非常简单,其基本转换方程的系数不随像幅、卫星号、接收站和接收时段而变化,也不依赖于特别试验区的条件,因而适合于大范围以至全球范围应用。而且这个新方法仍保持了KL变换、KT变换等方法的基本优点:降维和无明显信息损失。

二、地物的三个主要遥感特性

新方法要求首先分析实际地物的遥感特性。为此最理想的是使用具有明确物理意义的遥感数据。几年以前,作者曾试图用含有明确物理意义的遥感数据辐射值来进行土壤资源探测^[12]。近年来,国外已有遥感科学家建议使用有明确意义的物理量——辐射值或反射率代替传统使用的磁带数码,并建议制作辐射值或反射率数值磁带以代替传统的磁带^[13]。

但迄今为止通用的陆地卫星数据主要是CCT磁带数码。有时也使用图像扫描或测量密度值。但无论前者或后者,都只是一些相对数,而不是有明确意义的物理量。

在本研究中,我们选用辐射值这个物理量来分析地物的遥感特性。并将非物理量磁带数码值或图片密度值转换成了物理量辐射值。下面我们就利用换算出来的辐射值来分析地物的遥感特性。

设有一地物,它在陆地卫星4个波段中的辐射值分别为 R_4, R_5, R_6 和 R_7 。以各波段的某一波长值 $\lambda_4, \lambda_5, \lambda_6$ 和 λ_7 (通常是波段中心的波长值,也可为该波段的其它波长值,例如0.55, 0.65, 0.75和 $0.90^{1)}\mu\text{m}$)为横坐标值,以各波段相应的辐射值 R_4, R_5, R_6 和 R_7 ($\text{mW}/\text{cm}^2 \cdot \text{Sr}$)为纵坐标值,可以点绘出该地物在 λ -R平面中的一条曲线(实为折线)。我们称它为波段辐射曲线。

图1是一些代表性地物的波段辐射曲线。分析图1,我们可以得到如下的一些概念:

首先,各种地物的波段辐射曲线,在 λ -R平面中有些处于较高的位置,即它们的总辐射水平较高,如528号和485号;另一些则处于较低的位置,即它们的总辐射水平较低,如423号和39号。由此可见,地物的总辐射水平的高低是地物的第一个重要遥感特性(辐射特性)。

第二,各种地物的波段辐射曲线在 λ -R平面中有一些向左倾斜,即它们的红外辐射相对较强,可见光辐射相对较弱,如181号和150号;另一些则向右倾斜,即它们的红外辐射相对较弱,可见光辐射相对较强,如39号和423号。由此可以说,红外辐射与可见光辐射的强弱对比关系,或者说它们的辐射平衡关系是地物的第二个重要遥感特性。

第三,各种地物的波段辐射曲线从一个波段到另一个波段变化的方向和强度不一样。有些地物的波段辐射曲线,如206号和200号,从第4波段到第5波段是上升的,第5到第6波段是下降的,第6到第7波段又转为上升;有些地物的辐射曲线,如69号和150号,从第4波段到第5波段是下降的,第5到第6波段是上升的,第6到第7波段也是上升的;另一些地物的辐射曲线,如39号和423号,从第4波段经过第5、第6直到第7波

1) 选0.90而不选 $0.95\mu\text{m}$ 是为了将更多的信息集中到前三个新提取的变量中。

段,都是下降的。而且,各曲线上升或下降的速度(曲线的陡度)也不相同。结果,各种地

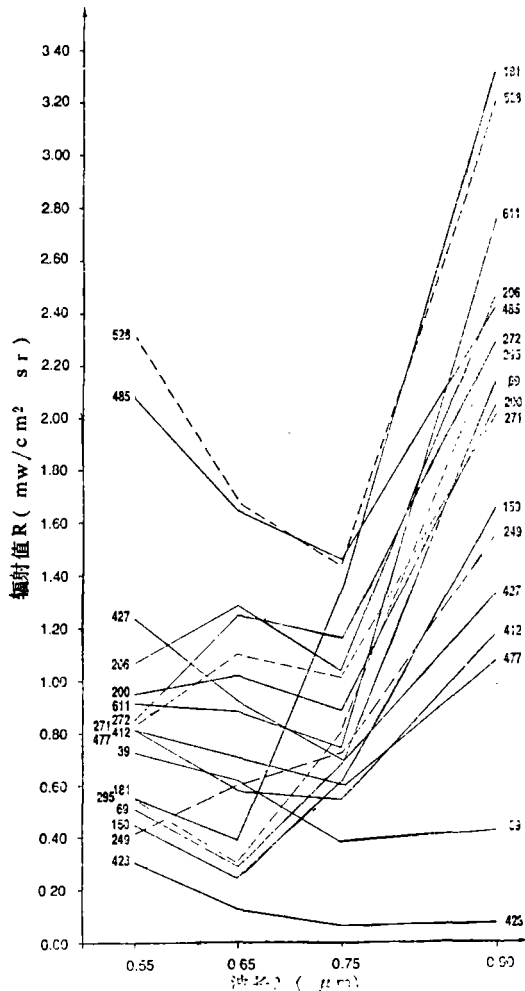


图1 世界范围内某些地物的波段辐射曲线
Fig. 1 The band radiation broken lines of some ground features in the world

181—热带森林,巴西,1977年7月; 528—云,中国新疆,1977年7月; 611—裸地,秘鲁皮乌腊与拉姆拜克,1980年12月; 206—侵蚀裸露土壤,澳大利亚,1973年5月; 485—雪,加拿大纽芬兰,1977年6月; 272—半荒漠沙丘状积沙地,苏丹,1979年10月; 295—苜蓿地,美国加利福尼亚,夏天; 69—阔叶林,意大利南部,1972年8月; 200—萨旺纳开阔地,澳大利亚,1973年5月; 271—半荒漠薄层土,苏丹,1979年10月; 150—针叶林,加拿大渥太华,1984年7月; 249—亮色的烧过的草地,澳大利亚,1975年4月; 427—冰,美国绿湾,1975年2月; 412—城市,加拿大渥太华,1984年7月; 477—沙丘,中国新疆,1976年9月; 39—浑水,中国南京,1979年6月; 423—较清的深水,美国绿湾,1975年2月。

样从数学上实现上述三个重要地物遥感特性的提取。

物的波段辐射曲线,有的呈上凸形,有的呈下凹形,有的呈平直形,而且其上凸下凹的程度也不相同。由此可以说,辐射随波段变化的方向和强度,是地物的第三个重要的遥感特性。

地物还有其它的辐射特性或遥感特性,但均不是重要的了(对MSS波段而言)。重要的只有上述三个。

由于地物的这三个遥感特性在遥感数据或遥感图像上有敏感的反应,因此,它们是我们分辨地物的主要依据。例如,我们用遥感图像或数据,分辨深色(或暗色)物体与浅色(或亮色)物体,就是根据地物总辐射水平的高低;划分水陆界线、水体与植物的界线或判断地面层水分含量的高低,就是根据地物可见光辐射与红外辐射的强弱对比关系或它们的辐射平衡关系;判断地面植物覆盖程度或地面裸露程度,就是根据地物总辐射水平的高低(特别是在目视判读的情况下),和地物辐射随波段变化方向和强度的不同(尤其是在计算机数据处理的情况下)。

反过来说,一个地物或它的某种性质,如果不在上述三个遥感特性方面,与其它地物或特性有所区别,那么该地物或该特性,就不能凭遥感图像或数据来识别。至少识别是十分困难的。

全色片或全波段数据在分辨地物方面的局限性,就是因为它们基本上只能反映地物的第一个遥感特性,即它们的总的辐射水平。

三、信息提取原理

现在从方法的逻辑与概念来讨论怎

我们已经看到,任一地物的 R_4, R_5, R_6 和 R_7 , 其物理含义都是辐射值, 只是分属于不同波段而已。如果我们能找到某种方法, 将每个波段的辐射值矢量, 即 R_4, R_5, R_6 和 R_7 , 都先分解为具有不同物理意义的几个部分, 然后再将所有波段中物理意义相同或相近的那些部分, 分别重新组合起来, 就有可能构成几个能反映前述地物主要遥感特性的新变量。

考察已有的主要数学方法后发现, 使用回归分析的方法, 可以达到此目的。

对地面上任一点的地物, 以其在卫星图像上对应点的 4 个波段的辐射值 R_4, R_5, R_6 和 R_7 为因变量, 以各波段对应的某一波长值 $\lambda_4, \lambda_5, \lambda_6$ 和 λ_7 (选定为 0.55, 0.65, 0.75 和 0.90 μm) 为自变量, 组成数据对, 作回归曲线, 我们就可将 4 个波段的辐射值 $R_i (i = 4, 5, 6, 7)$ 都分解为如图 2 所示的不同意义的三部分 l_i, b_i 和 $v_i (i = 4, 5, 6, 7)$ 。

研究发现, 各波段辐射值被这样分解而成的三部分: $l_4, l_5, l_6, l_7; b_4, b_5, b_6, b_7; v_4, v_5, v_6, v_7$ 具有如下的性质:

(1) l_i 的性质

数学性质 I: $l_4 = l_5 = l_6 = l_7$ (1)

我们用 L_0 来表示它们。即

$$L_0 = l_4 = l_5 = l_6 = l_7 \quad (2)$$

从几何意义上来说, 此 L_0 是回归曲线最低点的纵坐标。其单位为 $\text{mW}/\text{cm}^2 \cdot \text{Sr}$ 。

L_0 的物理意义是它能反映一个地物的总辐射水平的高低。即前述地物的第一个重要遥感特性。它是地物最一般也最重要的遥感特性。

(2) v_i 的性质

数学性质 II₁:

$$v_4 + v_6 = -(v_5 + v_7) \quad (3)$$

数学性质 II₂: v_4 和 v_6 的数值同符号; v_5 和 v_7 的数值同符号; v_5 和 v_7 的数值的符号与 v_4 和 v_6 的数值的符号相反。

数学性质 II₃: v_4, v_5, v_6 和 v_7 的数值(绝对值)的大小成固定的比例, 即

$$v_4 : v_5 : v_6 : v_7 = a : b : c : d \quad (4)$$

式中 a, b, c, d 是一些固定的常数。它们与地物无关, 不随地物各波段辐射值的大小而改变, 而是只与作波段辐射曲线时选定的坐标系的横坐标值有关。在 $\lambda_4, \lambda_5, \lambda_6, \lambda_7$ 分别选定为 0.55, 0.65, 0.75, 0.90 μm 的情况下,

$$a : b : c : d = 0.15 : 0.42 : 0.35 : 0.08 \quad (5)$$

数学性质 II₃ 的推论 I: 在 $\lambda_4, \lambda_5, \lambda_6, \lambda_7$ 选定的情况下, a, b, c, d 绝对值的大小顺序固定不变。例如在取 $\lambda_4 = 0.55, \lambda_5 = 0.65, \lambda_6 = 0.75, \lambda_7 > 0.85$ 的情况下, $a, b, c,$

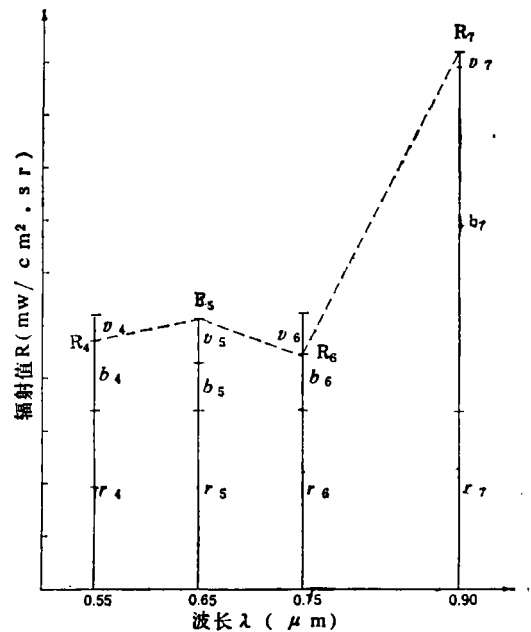


图 2 辐射值数据信息分解与重组合示意
Fig. 2 Sketch map of the discomposition and recombination of the radiance data

d 绝对值的大小顺序为:

$$|b| > |c| > |a| > |d| \quad (6)$$

数学性质 II₃ 的推论 II: 若 ν_5 为 0, 则 ν_7 必为 0, 且 ν_4 和 ν_6 亦必为 0; 若 ν_6 为 0, 则 ν_4 必为 0。

ν_4, ν_5, ν_6 和 ν_7 具有上述这些数学性质, 对信息提取是极端重要的。这意味着, 我们只需知道其中任何一个的数值, 便可确切地知道其它 3 个的数值。也就是说, 其中任何一个数值, 都包括了 4 个数值所包含的信息的全部。因此, 虽然从原有不同波段的 4 个辐射值分解出来的 $\nu_4, \nu_5, \nu_6, \nu_7$ 也有 4 个值, 但原则上说, 我们可以取其中任何一个作为新变量来代表它们 4 个。这大大有利于降低被处理的新变量的个数或数据的维数。

考虑到 ν_4, ν_5, ν_6 和 ν_7 的绝对值的相对大小, 我们可以选择其中绝对值较大的一个, 例如 ν_5 或 ν_6 来代表它们。

但, 考虑到 ν_4, ν_5, ν_6 和 ν_7 总的来说都是比较小的值, 且有正有负, 相加会互相抵消, 我们可以设想用它们的绝对值的和来代表它们。但由于 $\nu_4, \nu_5, \nu_6, \nu_7$ 数值的正负号包含有重要信息, 因此, 又应设法保留它们的正负号。也就是说, 用它们的绝对值的和来代表它们也不行, 得另想办法。研究发现, 我们可以定义一个特殊的和, 来达到既用它们的和又保留它们各自的正负号的目的, 即令

$$V_0 = -\nu_4 + \nu_5 - \nu_6 + \nu_7 \quad (7)$$

来作为新变量。

此式的设计考虑到了前述整个数学性质 II 及其推论。

从几何意义上说, V_0 是回归残差(即 ν_4, ν_5, ν_6 和 ν_7) 的特殊定义和式(7)。其单位亦是 $\text{mW}/\text{cm}^2 \cdot \text{Sr}$ 。

V_0 的物理意义是, 它能反映一个地物在整个卫星接收波段(MSS 波段)范围内, 从一个波段到另一个波段辐射变化的方向和强度, 即前述地物的第三个遥感特性。而这个特性又反映地面植被与裸地的对比关系, 或者说植物的覆盖度和生活强度的总情况。若 V_0 为负, 则地面植物密度较大, 植物生长旺盛; V_0 的负值的绝对值愈大, 则植物的总状况愈好。若 V_0 为正, 则地面较裸露; 正值愈大, 地面裸露程度愈大。因此, V_0 是地面植物生产量的良好指标。 V_0 还有其它作用, 此处不赘述。

(3) b_i 的性质

数学性质 III₁: b_i 是另外两个量 B_0 和 C_0 的函数。即

$$\left. \begin{aligned} b_4 &= (B_0 - 0.55)^2 C_0 \\ b_5 &= (B_0 - 0.65)^2 C_0 \\ b_6 &= (B_0 - 0.75)^2 C_0 \\ b_7 &= (B_0 - 0.90)^2 C_0 \end{aligned} \right\} \quad (8)$$

数学性质 III₂: b_i 的值主要取决于 B_0 , 且 C_0 可以表示为 B_0 的近似函数。

数学性质 III₁ 和 III₂ 的推论: b_i 的信息主要包含在数 B_0 中, 而 C_0 是可以忽略的。

由上述情形来看, 可以取 B_0 为另一个新变量。

从几何意义上来说, B_0 是回归曲线最低点的横坐标。其单位是 μm 。

B_0 的物理意义是它能反映一个地物的可见光辐射与红外辐射的对比关系, 或者说可

见光与红外的平衡关系,即前述地物的第二个重要遥感特性。这一平衡值 B_0 在其高值区反映地面水分的多少。 B_0 值愈大,水分状况愈好。当地面完全为水占据时, B_0 值最高。因此 B_0 是地面水体存在与水分含量的一个良好指标。在 B_0 的中低值区, B_0 亦可反映地面植物密度和生活强度的情况。即 B_0 值愈小,植物状况愈好。但这一关系不十分稳定。例如有些裸地也可以有很低的 B_0 值。由此可见, B_0 可作为地面植物状况的一个辅助指标,但不如 V_0 作为植物状况指标好。 B_0 还有其它的功能。此处亦不赘述。

关于 L_0 , B_0 和 V_0 的具体计算方法,作者将另文讨论。这里仅列出其最后的方程组:

$$\left. \begin{aligned} V_0 &= -0.457604R_4 + 1.28129R_5 - 1.06774R_6 + 0.195271R_7 \\ C_0 &= 19.3411R_4 - 14.1550R_5 - 21.5375R_6 + 13.0811R_7 \\ B_0 &= (15.3005R_4 - 9.8414R_5 - 15.9656R_6 + 8.4052R_7)/C_0 \\ L_0 &= 11.9112R_4 - 6.35144R_5 - 11.2071R_6 + 5.3179R_7 - C_0B_0^2 \end{aligned} \right\} \quad (9)$$

$$\left. \begin{aligned} V_0 &= v_{00} + v_{04}D_4 + v_{05}D_5 + v_{06}D_6 + v_{07}D_7 \\ C_0 &= c_{00} + c_{04}D_4 + c_{05}D_5 + c_{06}D_6 + c_{07}D_7 \\ B_0 &= (b_{00} + b_{04}D_4 + b_{05}D_5 + b_{06}D_6 + b_{07}D_7)/C_0 \\ L_0 &= l_{00} + l_{04}D_4 + l_{05}D_5 + l_{06}D_6 + l_{07}D_7 - C_0B_0^2 \end{aligned} \right\} \quad (10)$$

$$\left. \begin{aligned} L &= 245L_0 - 56L_0^2 \\ B &= 486 - 2950B_0 + 5428B_0^2 - 2714B_0^3 \\ V &= 128 - 320V_0 + 300V_0^3 \end{aligned} \right\} \quad (11)$$

方程组(9)适用于陆地卫星数据已经是辐射值(R_4, R_5, R_6, R_7 , 单位是 $\text{mW}/\text{cm}^2 \cdot \text{Sr}$)的情形。 L_0 和 V_0 的单位仍是 $\text{mW}/\text{cm}^2 \cdot \text{Sr}$, B_0 的单位仍是 μm 。方程组的系数值是固定的,不依卫星号、接收站、接收时段、像幅或数据分辨率而变更,更不依赖于特定的地域条件。

式中的 C_0 是第 4 个新变量。但 C_0 的信息基本上已包括在 L_0, B_0, V_0 之中,主要在 B_0 之中。实际计算证明, L_0, B_0, V_0 三个变量包含的信息占原有四个变量 R_4, R_5, R_6, R_7 总信息量的 97% 以上。故实际图像处理中仅使用 L_0, B_0, V_0 就够了。这里 C_0 仅作为计算 B_0 和 L_0 的一个桥梁。一般也不必输出。

方程组(10)适用于现在通用的陆地卫星 CCT 磁带。式中的 $D_i (i = 4, 5, 6, 7)$ 是波段 i 的磁带数码值, $v_{0i}, c_{0i}, b_{0i}, l_{0i} (i = 0, 4, 5, 6, 7)$ 都是常数。将各陆地卫星、各接收站和各接收时段的各波段的最高辐射值、最低辐射值、最高 D_i 数码值等代入相应的辐射定标计算公式,即可算得这些常数的值。这些常数已将辐射值转换计算包含在内。

对于来自美国和加拿大的各个陆地卫星、各个接收站、接收时段和各种数据分辨率的磁带数据,作者已计算出方程组(10)的系数 40 多套,并制成系数表供查阅。

方程组(11)是将 L_0, B_0 和 V_0 进行拉伸或压缩同时使其最后输出值 L, B, V 仍为 0—255 的数码以供计算机处理的公式。此式的设计,首先是考虑各种地物的 L_0, B_0 和 V_0 在各自数轴上的分布性质和分布范围;其次是考虑为了降低图像处理成本应使计算越简单越好。

实际上,包括辐射值换算、信息提取、数据拉伸-压缩以及转变成 0—255 的数码等全部运算,只需使用 7 个简单的算术式,即方程组(10)和方程组(11)就够了。此方法计算之

简单,用机时之少,可想而知。

对于卫星像片扫描密度值,按作者推荐的方法和公式^[12,14]算得辐射值后,代入方程组(9)和(11)即可。

四、新变量的功能

前面已经简单提到 L_0 , B_0 , V_0 (L , B , V 的前身) 的物理意义。这里再对 L , B , V 的意义作进一步的说明。

让我们来看图 3 中若干组地物的波段辐射曲线。

图 3a 说明新变量 L 的功能。第一组曲线(实线)中的 441 号地物(美国得克萨斯的银叶向日葵地,卫星 3, 1977 年 6 月 2 日)和 139 号地物(加拿大渥太华的森林和草地混合地,卫星 5, 1984 年 7 月 1 日)具有相同的 B 值和几乎相同的 V 值。二者几乎为形状相同的平行曲线。这意味着二者具有相同的可见-红外辐射平衡和几乎相同的波段辐射变化方向和速度。但是, 441 号的 L 值为 107, 而 139 号为 80。441 号曲线在坐标平面中处于较高的位置。这表明 L 值大的地物其总辐射水平较高。第二组曲线(虚线)的情形也是一样(其中 25 号地物为中国南京的河流冲积物沙土地,卫星 3, 1979 年 6 月 13 日; 215 号地物为中国新疆的红柳林盐土地,卫星 1, 1972 年 11 月 15 日)。我们将 L 命名为地物的总辐射水平值。

图 3b 说明新变量 B 的功能。第一组曲线(实线)中的 343 号地物(澳大利亚科巴尔科里巴的灌木林,卫星 2, 1975 年 8 月 8 日)和 122 号地物(加拿大渥太华的沙地和沙砾地,卫星 5, 1984 年 7 月 1 日)具有相等的 L 值和几乎相等的 V 值,二者在坐标平面中总的来说处于同样高的水平,且两曲线的形状也很相似。这表明二者具有相同的总辐射水平和几乎相同的波段辐射变化的方向和强度。但是, 343 号的 B 值小于 122 号的 B 值。这使得 343 号曲线相对地向左倾斜,即它的红外辐射相对较强,可见-红外辐射平衡点移向可见光的一端;而 122 号曲线相对地向右倾斜,即它的红外辐射相对较弱(与 343 号比),可见-红外辐射平衡点移向红外光一端。第二组曲线(虚线)说明了类似的情形(其中 154 号地物为加拿大渥太华河中的浅水段或急流段,卫星 5, 1984 年 7 月 1 日; 194 号地物为菲律宾拉贡德贝泻湖湖水,卫星 2, 1976 年 4 月)。我们将 B 命名为地物的可见-红外辐射平衡值。

图 3c 说明新变量 V 的功能。第一组曲线(实线)中的 225 号地物(中国新疆的龟裂性土或龟裂状棕漠土,卫星 1, 1972 年 11 月 15 日)和 23 号地物(中国南京长江沙洲耕地,卫星 3, 1979 年 6 月 13 日)具有完全相同的 L 值和 B 值,二者在坐标平面中总的说来处于同样高的辐射水平和同样的可见-红外辐射平衡状态。但 225 号的 V 值小于 23 号的 V 值,二者具有不同的曲线形状和波段辐射变化规律: 225 号在可见光波段范围内明显地向上拱起,在红外波段则下凹较剧,即它由第 4 到第 5 波段辐射变化微小,由第 5 到第 6 波段辐射呈较急剧下降势,由第 6 到第 7 波段辐射呈较急剧上升势;而 23 号在可见光波段明显下凹,在红外区域下凹较和缓(与 225 号比),即它从第 4 到第 5 波段辐射呈明显的下降,从第 5 到第 6 波段辐射继续下降,但比 225 号和缓,从第 6 到第 7 波段辐射也上升,

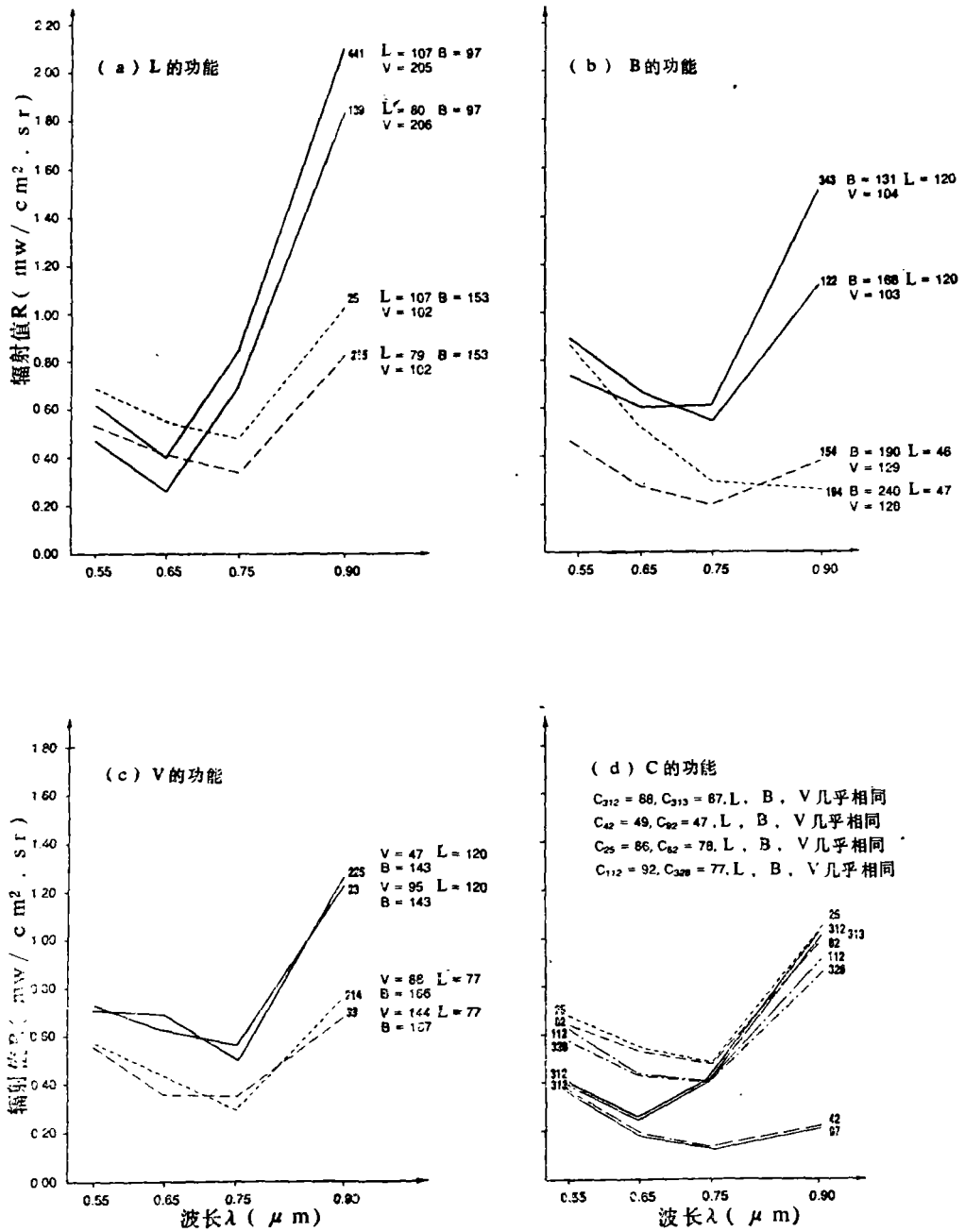


图 3 新变量 L, B, V 的功能

Fig. 3 Function of the L, B, V

但较 225 号和缓。这种差别意味着 225 号地面裸露程度比 23 号高。第二组曲线(虚线)表明了类似的情形(其中 214 号地物为中国新疆的典型盐土, 卫星 1, 1972 年 11 月 15 日; 38 号地物为中国南京的鱼塘, 卫星 3, 1979 年 6 月 13 日)。我们将 V 命名为地物波段

辐射变化方向和强度(速度)值或波段辐射变化向量值。

图 3d 表明第 4 个变量 C 的功能。在它的四组曲线的每一组曲线中,两个地物具有几乎相同的 L 值、B 值和 V 值。因而两个地物处于几乎同样高的总辐射水平,几乎同样的可见-红外辐射平衡状态和具有几乎同样的波段辐射变化方向和强度。但是, C 值却有两种

情况。第一种情况是两个地物也具有相同的或几乎相同的 C 值,如 42 号和 97 号, 312 号和 313 号。这时它们的两条曲线几乎没有差别。这就是说,只要两个地物的 L, B, V 值相同,它们的两条曲线就相同。或者说, C 的信息完全包含在 L, B, V 之中,因而 C 是多余的量或裕量,是可以忽略不计的。第二种情况是两个地物的 C 值有明显的不同。例如 112 号和 328 号, 25 号和 82 号。即使如此,它们的曲线也只有非常微小的差别。这意味着,包含在 C 中的独立信息是非常少的,不重要的。例如分类时的类内差别等。这时 C 同样是可以忽略的。

仔细观察 112 号和 328 号曲线,或 25 号和 82 号曲线,我们可以看到 C 值较大的曲线(如 112 号)比 C 值小的曲线(如 328 号)两端上翘得高一些。因此,变量 C 可以称之为高辐射波段(通常是 4 和 7 波段,有时是其它)与低辐射波段(通常是 5 和 6 波段,有时是其它)的辐射对比值。

将图 1 中的各种地物用本方法进行信息提取,得到它们的 L, B, V 值,绘制成曲线(折线),得图 4。比较图 1 和图 4,可以明显地看出新变量的优越性(含

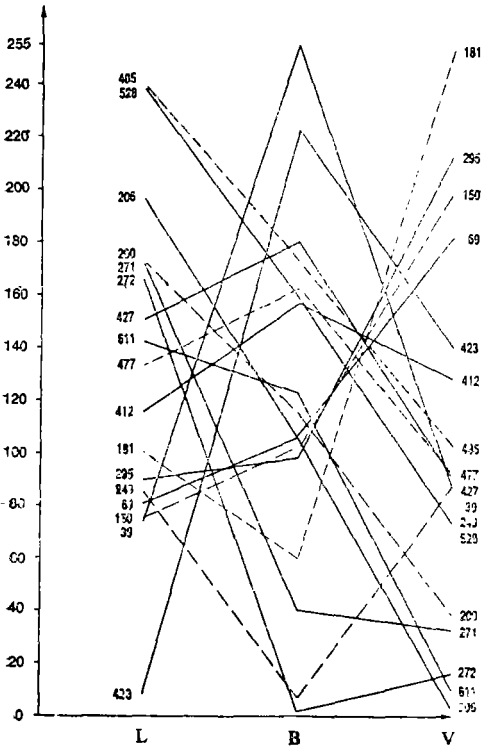


图 4 图 1 中所示地物的 LBV 曲线

Fig. 4 The LBV broken lines of the same ground features as in Fig. 1.

181. 热带森林, 295. 苜蓿地, 150. 针叶林, 69. 阔叶林, 423. 水, 412. 城市, 485. 雪, 477. 沙丘, 427. 冰, 39. 混水, 249. 亮色的烧过的草草地, 528. 云, 200. 萨王纳开阔地, 271. 半荒漠薄层土, 272. 半荒漠沙丘积沙地, 611. 裸地, 206. 侵蚀裸露土壤

义明确性,可分性,降维等)。

五、新方法的意义和问题

与使用陆地卫星图像原有 MSS 4 个波段的数据相比,使用本方法提取出来的新信息(新变量)有如下优点:(1)数据维数降低,一般只需使用 3 个变量。这将使计算时间缩短,图像处理成本降低。(2)变量物理含义明确清晰。(3)分辨地物的能力强。(4)提供了世界范围内不同地区、不同时间、不同卫星、不同像幅、不同数据分辨率的图像及其处

理结果定量比较的可能性。

与 KL 变换相比,这种新的信息提取方法有下述优点:(1) 计算非常简单,这将大大降低图像处理的成本。(2) 转换方程的系数是已知的,不随地域条件和所用像幅或其子区而改变,不需要在每次图像处理时都重新计算转换方程的系数。(3) 新变量物理含义明确。(4) 分辨地物的能力强。(5) 所得新图像及其处理结果在相当程度上有世界范围内的数值可比性。

与 KT 变换相比,其优点是:(1) 转换方程的系数不依赖于特定地域条件和原始试验点的选择,且其基本方程的系数不因卫星不同而变化。(2) 具有广阔范围直至世界范围内的适用性。(3) 新变量物理含义明确。(4) 分辨地物的能力强。(5) 所得新图像及其处理结果在相当程度上具有世界范围的数值可比性。

本方法提取出来的第三个新变量 V,其物理含义独特,分辨地物能力很强。无论是 KL 变换,还是 KT 变换,或是其它变换,都没有产生出这样的类似变量来。本方法提取出来的第二个新变量 B,在对于水分和水体以及某些其它地物的分辨方面,也比较独特和有效,为其它方法所不及。本方法提取出来的第一个新变量 L,比较接近于 KL 变换和 KT 变换提取出来的第一个变量(常被叫做亮度),但所含信息更加明确、单纯。

新变量 L, B, V 用于彩色合成时,比原来 4 个波段数据的任何三个波段的组合都好。因为这三个新变量包含的信息比原波段任何三个的组合包含的信息都多。它们正好能充分利用颜色的三原性(红、绿、蓝)。不仅如此,由于 L, B, V 的特定性质,它们还能合成出一种近于自然色彩的“真”彩色合成图像来。例如,赋予 V 值以绿色,赋予 B 值以蓝色,赋予 L 值以红色,并给定一个滤色界限值,使大于某 L 值(相当于云和雪的下限值)的一切 L 值和对应的 B、V 值均为 255,那末植被将为绿色(V 值大),水体将为蓝色(B 值大),裸地将为红色(L 值大),云和雪将为白色。其它地物也将具有比较自然的中间彩色。

新变量用于密度分割,可得到有明确特征地物分级的结果。如植物密度等级、水分含

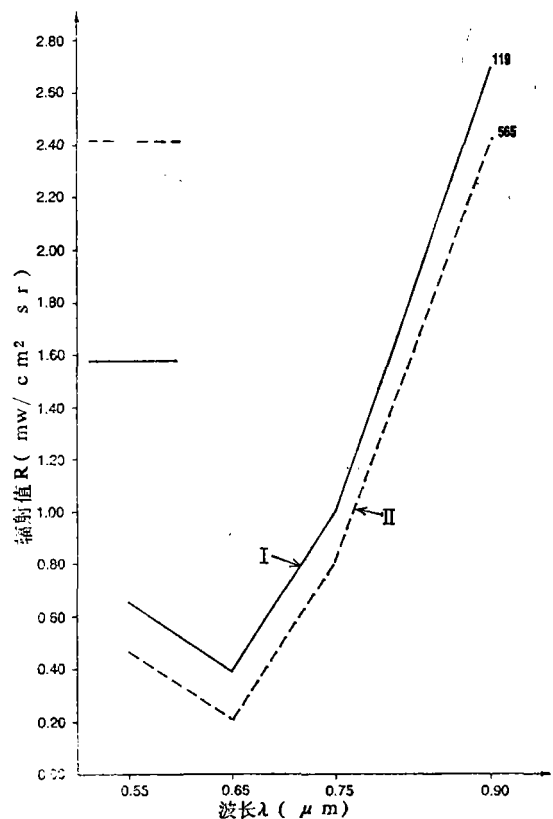


图 5 上空有不可分辨的云(折线 I, $V = 226$, $B = 95$, $L = 111$)的很密植被和没有不可分辨的云(折线 II, $V = 226$, $B = 92$, $L = 77$)的很密植被

Fig. 5 Very dense vegetation with and without unresolved cloud

量等级等。

此外,目前困扰着许多遥感科学家的小于像元或不可分辨的云,若同时使用 L 和 V 则可以识别(见图 5)。

作者已经将本方法应用到世界范围内的资料,取得了良好结果¹⁾。

本方法尚有一些问题待研究解决。

文中所说的世界范围内的可比性是指辐射值及其转换数据。但太阳高度角和大气传输,对辐射值会产生一定的或相当的影响。

太阳高度角的问题,我认为可以将辐射值转换成反射率以消除之。本研究中,辐射值的比较是采取高太阳高度角(夏天及其前后)和低太阳高度角(冬天及其前后)两种情况来处理的。

大气传输问题,世界上虽已提出了种种消除其影响的模式,但由于缺乏足够的地面或空中的其它同步观测数据,故实行起来颇难。作者建议分较高透明度大气(基本为晴天)和较低透明度大气(透明度较低,但仍能接收数据)两种情况,分别比较。本文亦是这样处理的。就多数实际工作而言,这样处理是可行的,数据一般也是可以比较的。因为大家一般并不要求理论上的严格准确性,而且在实践中总是选用较高透明度大气下的图像,个别也使用较低透明度大气下的图像,而不用透明度很低或不透明大气下的图像。

最后,关于 TM 图像,本方法的逻辑和概念仍然适用,但对具体的新变量需作一些适当的新规定,因之计算方法也应随之作一些变更。

参 考 文 献

- [1] Boyd, R. K., J. O. Brumfield, and W. J. Campbell, 1983. A Comparison of the Usefulness of Canonical Analysis, Principal Components Analysis, and Band Selection for Extraction of Features from TMS Data for Landcover Analysis. Proceedings of the Seventeenth International Symposium on Remote Sensing of Environment. Vol. III, 9—13, Ann Arbor, Michigan. 1983.
- [2] Giddings, L. E., M. Soto, M. J. Angulo, H. Larios and A. M. Ramirez, The Use of Landsat Data in Mapping Tropical Vegetation. Proceedings of the Fourteenth International Symposium on Remote Sensing of Environment Volume III 23—30 San Jose, Costa Rica, 1980.
- [3] Gray, T. I., and D. G. McGray, The Environmental Vegetation Index, a Tool Potentially Useful for Arid Land Management. AgRISTARS Report, No. EW-NI-04076, JSC-17132, 1981.
- [4] Howard, J. A., M. Lorenzini and A. Zandonella, A procedure for the Extraction of Vegetation Information from NOAA AVHRR Data. Proceedings of the Eighteenth International Symposium on Remote Sensing of Environment. Paris, France, October 1—5, 1984.
- [5] Jackson, R. D., Spectral Indices in n-Space. Remote Sensing of Environment. 13: 409—421, 1983.
- [6] Kauth, R. J. and G. S. Thomas, The Tasseled cap— A Graphic Description of the Spectral-temporal Development of Agricultural Crops As Seen by Landsat. Proceedings of the Symposium on Machine Processing of Remotely Sensed Data, Purdue University, West Lafayette Indiana, 1976.
- [7] Pendock, N., M. Sears and A. A. de Gasparis, Reducing the Spectral Dimension of Remotely Sensed Data and The Effect on Information Content. Proceedings of the Seventeenth International Symposium on Remote Sensing of Environment. Ann Arbor, Michigan, May 9—13, 1983.
- [8] Santisteban, A., and L. Munoz, 1978. Principal Components of a Multispectral Image: Application to a Geological Problem. IBM J. Res., Develop., Vol. 22, No. 5, September 1978.
- [9] Taylor, M. M., Principal Components Colour Display of ERTS Imagery: Proceedings of the Second Cana-

1) 曾志远,新的卫星数据信息提取方法在世界范围内的应用,第五届全国遥感技术交流会论文集,第 154—159 页,1988 年。

- dian Symposium on Remote Sensing, University of Guelph, Canada, April 29-May 1, 1974.
- [10] Wiegand, C. L., and A. J. Richardson, Comparisons among a New Soil Index And Other Two- and Four-Dimensional Vegetation Indices. Technical Papers of the American Society of Photogrammetry, ACSM-ASP Convention, March 14—20, Denver, Colorado, 1982.
- [11] Schowengerdt, R. A., Techniques for Image Processing and Classification in Remote Sensing. Academic Press, INC. pp. 167—170. 1983.
- [12] 曾志远, 卫星图像密度-辐射量算与土壤资源探测, 自然资源, (4), 1981。
- [13] Robinove, C. J., 1982. Computation with Physical Values from Landsat Digital Data. Photogrammetric Engineering and Remote Sensing, Vol. 48, No. 5, May 1982.
- [14] 曾志远, 计算机分类及其结果的光谱学和地理学分析(卫星图像土壤类型自动识别与制图的研究), 土壤学报, 21(2), 1984。

A New Method for Extracting Information from Landsat Data

Zeng Zhiyuan

(Institute of Soil Science, Academia Sinica)

Abstract

The digital counts or the scanned densities of the Landsat MSS image were transformed into absolute radiance values. Each radiance of the four bands were decomposed into three parts with different significances and then the parts with the similar significance from the four bands were recombined. As a result, we obtained three new variables. Which are named respectively the general radiance level (L), the visible-infrared (radiance) balance (B) and the band radiance) variation vector (V). The L is powerful in monitoring radiation intensities of ground features; the B is powerful in mainly monitoring the water, the moisture or wetness of ground features and in differentiating the cloud, snow and ice among them; the V is powerful in monitoring density of plant cover and its vigor and hence is an excellent indicator of vegetation conditions. Those three variables contain nearly all the information of the original four bands, so we need only manipulating the three instead of four variables. The algorithm is based on the regression analysis and is very simple. Lower dimension and simpler calculation will lead to time and money saving.

Using this method, we can obtain "real" (not false) color composites, which are real three-coloured and near to the natural color. And we can obtain significant, corresponding to the ground feature characteristics, density slicing images.

The methodology described above was applied to the data on hand from China and Canada and other data found in literatures from Asia, Europe, Africa, Australia, Oceania and North and South America.

Key words Image processing Information extraction Feature extraction LBV transformation KL transformation Kauth-Thomas transformation Digital count Density value Radiance value Radiation calibration Radiance value calculation Color composition Density slicing Regression analysis