

# TM 图像的信息量分析及特征信息提取的研究

陆灯盛 游先祥 崔赛华

(北京林业大学)

1990年3月10日收稿

## 摘 要

图像信息量分析是图像处理的基础,为此,本文研究了三个不同植被覆盖类型区,即多林区(森林覆盖在40%以上)、一般林地分布的丘陵区(森林覆盖10—30%)和农田为主的丘陵与平原区的图像信息量。分析同一地区冬夏两季的图像信息特征后得知,红外波段的信息量高于可见光波段,其中信息量最大的是 TM5 波段,最小的是 TM2 波段。同时对不同情况下波段间的相关性、均值和标准差等统计特征值也进行了分析。据此就图像增强、信息特征提取方法,如主成分分析、缨帽变换(KT变换)、比值等方法以及波段组合等进行了系统研究,并就其实用条件进行了探讨和评价。

**关键词** 信息量分析 统计特征值分析 图像增强 信息特征提取 波段组合

## 一、TM 图像信息特征分析

设一幅图像有  $S_1, S_2, \dots, S_q$  共  $q$  种灰度值,其出现的概率分别为  $P_1, P_2, \dots, P_q$ , 则每一灰阶所具有的信息量分别为  $\log_2(1/P_1), \log_2(1/P_2), \dots, \log_2(1/P_q)$ 。因此,每幅卫星图像的单波段信息量为

$$H = - \sum_{i=1}^{55} P_i \log_2(P_i) \quad (1)$$

式中  $P_i$  为第  $i$  灰阶出现的频率,  $H$  为单波段影像的信息量<sup>[1]</sup>。

在图像信息量计算中,选择了不同季节、不同林区、少林区的 TM 不同波段进行了分析。数据分析表明:

(1) 信息量的大小与波段有关。表1表明,尽管地区不同,森林覆盖不同,一般红外

表1 不同时相、不同森林覆盖各波段信息量表

Table 1 Information content table in various bands of different temporals and forest covers

项目 \ 波段		TM1	TM2	TM3	TM4	TM5	TM6	TM7	$\Sigma$
		时相							
	夏季(山西)	4.474	4.069	4.384	5.888	6.283	5.519	5.200	35.817
	冬季(山西)	4.396	3.968	4.704	5.091	6.382	5.559	5.485	35.636
不同森林覆盖	少林(宁夏)	5.514	5.112	5.774	5.500	6.408		6.021	34.329*
	多林(宁夏)	4.667	4.501	4.634	6.349	6.406		5.210	31.77*

\* 少 TM6 波段

波段(包括近、中、热红外)的信息量大于可见光波段的信息量,且TM5信息量最大, TM2信息量最小。然而随植被覆盖度大小不同,其大小排列次序有一定的变化,但变化不大。

(2) 图像上亮度值的分布,在可见光区间为低亮度,红外波段为高亮度区,但 TM7 由于水的影响亦为低亮度区(表 2)。

表 2 不同时相、不同森林覆盖各波段统计特征值表

Table 2 Table of statistic characteristic value in various bands of different temporals and different forest covers

项目 \ 波段		TM1	TM2	TM3	TM4	TM5	TM6	TM7	
时 相	冬 季	亮度值范围	35—79	10—45	8—62	6—64	3—120	46—117	0—69
		均值	46.28	18.21	20.42	26.39	46.29	77.70	22.04
		标准差	5.83	4.54	7.51	8.09	21.39	10.5	12.09
	夏 季	亮度值范围	53—124	19—74	14—99	39—143	25—172	116—206	6—106
		均值	65.25	29.32	26.73	81.15	75.65	139.25	27.34
		标准差	10.04	7.49	12.73	13.81	22.88	15.03	16.75
不 同 森 林 覆 盖	多 林	亮度级	80	55	77	139	154		96
		均值	57.0	26.8	23.9	86.6	74.4		25.5
		标准差	9.18	6.74	11.25	18.12	21.31		14.91
	少 林	亮度级	86	64	103	126	204		126
		均值	96.9	51.4	64.6	76.0	126.3		74.5
		标准差	11.11	8.32	13.42	11.77	22.40		16.47

(3) 标准差的大小,说明图像中各地类亮度之差异,这种差异越大,分类效果越好。计算结果表明,通常也是红外波段的标准差大于可见光的标准差。仅在少林区的图像上, TM4 < TM3 (见表 2)。表 2 还表明,夏季的均值和标准差均大于冬季的均值和标准差。

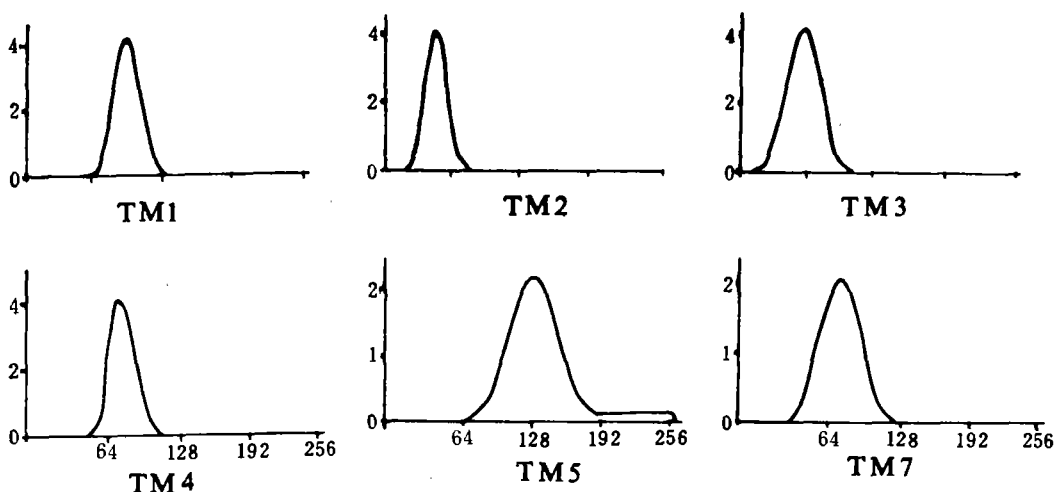


图 1 少林地区 TM 影像各波段直方图

Fig. 1 Histogram of different bands TM sparse forest area

在多林区和少林区由于水分条件相差大(多林区湿润,少林区干燥),在 TM5, TM7 波段因水对光谱吸收的影响,其均值和标准差多林区均小于少林区。

(4) 直方图直观地说明了(图 1,2),多林区在可见光波段亮度分布很窄,少林区则较宽,在红外波段则相反,而多林区 TM7 由于水的吸收作用,亮度分布则较少林区的 TM7 窄。

(5) 相关矩阵计算结果表明(表 3,4),可见光 TM1, TM2, TM3 各波段间相关系数大,在少林区和冬季图像中,各波段相关系数变动不大,在多林区和夏季图像中,则各波段间相关系数变动较大。特别是 TM4 在多林区和夏季图像上与所有波段相关都不紧密。国内外研究证明,在植被信息提取中, TM4 波段是不可缺少的波段。

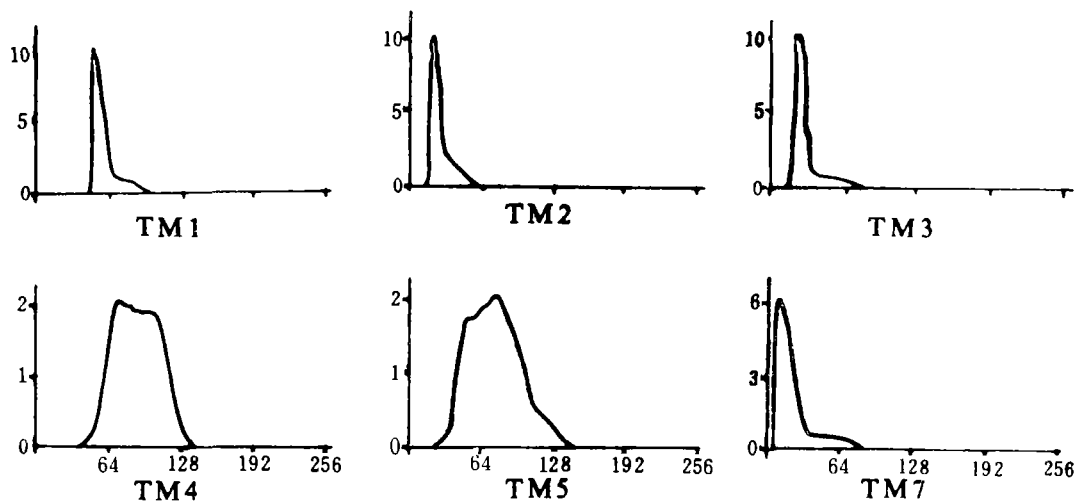


图 2 多林地区 TM 影像各波段直方图

Fig. 2 Histogram of different TM bands in multi-forest area

表 3 夏季/冬季 TM 各波段相关矩阵

Table 3 The cross correlation of TM different bands in summer/winter

	TM1	TM2	TM3	TM4	TM5	TM6	TM7
TM1	1.000/1.000	0.983/0.971	0.987/0.974	-0.008/0.814	0.908/0.905	0.894/0.652	0.978/0.950
TM2		1.000/1.000	0.986/0.984	0.086/0.845	0.929/0.891	0.887/0.640	0.979/0.941
TM3			1.000/1.000	-0.024/0.841	0.896/0.917	0.888/0.643	0.982/0.958
TM4				1.000/1.000	0.300/0.838	-0.005/0.710	0.047/0.823
TM5					1.000/1.000	0.850/0.761	0.944/0.976
TM6						1.000/1.000	0.902/0.739
TM7							1.000/1.000

表 4 少林区 TM 数据 KL 变换统计值

Table 4 statistical value of KL transformation of TM data in sparse forest area

波 段	TM1	TM2	TM3	TM4	TM5	TM7
均值	96.9	51.4	64.7	76.0	126.3	74.5
标准差	1.105	8.323	13.424	11.774	22.403	16.474
特征值 $\lambda_i$	5.448	0.306	0.210	0.024	0.012	0.001
方差百分比(%)	90.8	5.0	3.6	0.4	0.0	0.1
累计贡献率(%)	90.8	95.8	99.4	99.8	99.9	100.0

## 二、信息特征提取方法的研究

### 1. 图像增强技术

图像增强实质上是有选择地加强图像中某些信息,同时削弱某些不需要的信息。实验应用了局部拉伸,线性及分段线性拉伸,对数、指数拉伸,高斯拉伸,Wallis 拉伸,直方图均衡化,直方图归一化,直方图调整等图像增强方法,分别对少林区和多林区的 TM 图像进行了试验。结果表明,在分析地物波谱特征的基础上,用局部拉伸或分段线性拉伸,可方便而灵活地突出所需要的某些地物信息;如果要全面地反映地物信息,且直方图是近似正态分布时,选用 Wallis 拉伸、直方图归一化处理效果较好;而对于多林区直方图严重偏斜的图像,为了突出植被信息,可以采用对数拉伸处理,以减少信息的损失。

### 2. KL 变换及 KT 变换

KL 变换是在统计特征基础上的多维正交线性变换,通过变换可消除原有图像间的相关性,把原来各波段图像的有用信息集中在数目尽可能少的新组分图像中,从而起到信息提取和数据压缩的目的。经 KL 变换后,第一组分取得方差的绝大部分信息,占 90.8%,其余组分的方差依次减少;而且前 3 个组分累计贡献率达 99.4%(见表 4)。各组分图像之间互不相关,因而通过假彩色合成可充分显示不同地物之间的差异。经 KL 变换的 3 个组分图像中,乔、灌、草、坡耕地、裸地、水体等地类清晰可辨,而且还消除了阴影的干扰,图像的整体判读性能较好。

分析特征向量可见,第一组分是原来各波段的加权和,其加权值与该波段的标准差大小成正比,其中以中红外(TM5 和 TM7)提供的信息量最多,反映了地物总的辐射差异;第二组分是 TM4 和 TM5 波段与其余波段的差值图像,其中以 TM4 提供的信息量居多,较好地反映了植被信息的特点;第三组分则是中红外波段与其余波段的差值,主要包含相关程度较高的波段间的信息差异;其余组分的信息量很小,主要是噪声。

KT 变换把景物的光谱特征与自然景观属性联系起来。通过变换,可获得 6 项特征,其中前 3 项给出了明确的物理——景观含义,分别定义为亮度、绿度和湿度,其余几个特征则是前 3 个特征遗漏的信息及噪声(表 5)<sup>[2,3]</sup>。

比较 KT 变换系数及 KL 变换的特征向量可见,两种变换方法的前 3 项特征是一致的,只是其系数不同。然而,从实际计算考虑,KT 变换相对地简单些,而且更切合实际。表 6 是 KL 变换与 KT 变换的结果比较。

可见, KL 变换和 KT 变换都具有较强的数据压缩能力, 把信息量绝大部分集中在前 3 个特征图像中, 其方差累计贡献率达 97% 以上。从第一组分的标准差和方差百分比可见, KL 变换的效率比 KT 变换的高。

上述已对 TM 数据的 KL 变换作了分析, 那么, 对比值图像的 KL 变换, 效果如何呢? 表 7 以西吉县不同图像子集的 KL 变换结果比较回答了这个问题。

表 5 TM 数据的 KT 变换系数

Table 5 KT transformation coefficient of TM data

波段 特征图像	TM1	TM2	TM3	TM4	TM5	TM7
亮度	0.3037	0.2793	0.4743	0.5585	0.5082	0.1863
绿度	-0.2848	-0.2435	-0.5436	0.7243	0.0840	-0.1800
湿度	0.1509	0.1973	0.3279	0.3406	-0.7112	-0.4572
第四特征	-0.8242	0.0849	0.4392	-0.0580	0.2012	-0.2768
第五特征	-0.3286	0.0549	0.1075	0.1855	-0.4357	0.8085
第六特征	0.1084	-0.9022	0.4120	0.0573	-0.0251	0.0238

表 6 KL 变换和 KT 变换的结果比较

Table 6 Comparison of result between KL transformation and KT transformation

波段	TM1	TM2	TM3	TM4	TM5	TM7
标准差	11.905	8.323	13.424	11.774	22.403	16.474
特征图像	PC1	PC2	PC3	PC4	PC5	PC6
KL 变换标准差	34.153	8.089	6.673	2.257	1.562	0.864
方差百分比(%)	90.8	5.0	3.6	0.4	0.1	0.1
KT 变换标准差	30.990	9.895	13.942	3.633	4.322	1.032
方差百分比(%)	74.7	7.6	15.1	1.0	1.5	0.1

可见, 不同图像子集经 KL 变换后其特征图像的信息量是不同的。其中比值图像子集变换后的特征图像标准差最大, 所包含的信息量也最多; 其次是混合图像子集。从信息量及图像的判读性能考虑, 采用比值图像或混合图像子集的 KL 变换效果要比原始 TM 波段的变换效果要好。

### 3. 比值变换

比值变换被认为是可以消除大气影响、地形因子影响的一个重要方法。在植被信息提取中, 主要是应用近红外的高反射与可见光波段中叶绿素吸收区间的大反差, 通过比值使其差异更突出。研究表明, 比值的前提是各波段的直方图分布和相关系数的大小(图 3 3a, 3b, 4a, 4b)。图 3a, 3b 仅有极少重叠区, 这种比值只会出现零值, 比值效果极差; 而图 4a, 4b 具有一定重叠, 其比值效果较好, 如既具有重叠又相关系数小, 则表明这两两对应波段的比值具有较好的效果。

根据上述讨论, 在分析地物波谱特性基础上, 对比值处理方案作了大量的试验。结果

表 7 不同方案的 KL 变换特征图象的标准差比较  
Table 7 Standard-deviation comparison of KL transformation feature image between different plans

波段	TM1	TM2	TM3	TM4	TM5	TM7	
变换前	11.105	8.323	13.424	11.774	22.403	16.474	
变换后	34.153	8.089	6.673	2.257	1.562	0.864	
比值图像	4/3	5/4	7/2	(2×4)/3	(3×4)/2	(4×5)/3	(5×7)/4
变换前	49.286	54.974	45.535	42.876	44.579	47.815	45.025
变换后	96.604	76.633	36.014	15.681	10.100	8.265	3.944
混合图像	TM3	TM7	绿度	4/3	5/4	(5×7)/4	
变换前	36.259	38.680	37.417	49.286	54.974	45.025	
变换后	88.119	44.210	30.278	8.026	7.159	2.750	

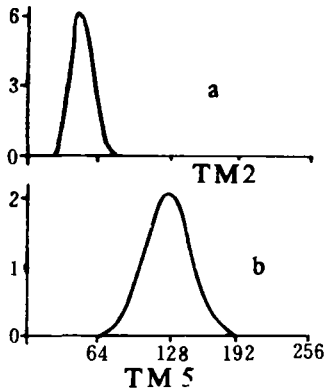


图 3 图 3a, 3b 仅有极少重叠的两个波段

Fig. 3 Two Bands of histogram  
a tiny minority overlap

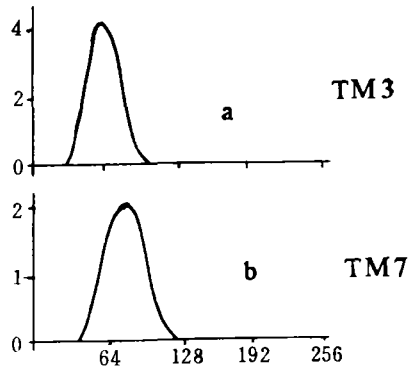


图 4 图 4a, 4b 为具有一定重叠的两个波段

Fig. 4 Two Bands of histogram  
certain overlap

表明,较好的波段比值是:(1)反映植被信息较好的单波段,如 TM4;(2)非线性波段比,如 $(4 \times 5)/3$ , $(3 \times 4)/2$ ;(3)波段间相关程度低的波段比,如 $4/3$ , $5/4$ , $7/2$ ;(4)反映植物生物量的波段比,如归一化差值指数。另外,差值指数在地形起伏不大时效果较好,但在地形起伏大,坡陡阴影严重时差值指数效果不好,而采用非线性波段比效果良好。对于单波段比值,发现相关性大的波段之比,如 TM1/TM2 等,其效果不理想。在组合图像中,如 3 个波段都由比值图像构成,在夏季时,因灌、草、农作物易误分,植被覆盖常增加 10—20%。

#### 4. 波段图像的组合分析<sup>[4]</sup>

TM 图像共有 7 个波段,由这 7 个波段图像经过各种增强处理、比值处理、变换处理等可以派生出许多新的独立的图像。在这众多的图像中,如何选取信息量丰富、判读效果最好的组合,目前采用的方法有两种。

一种是以研究地区的地面或航空波谱测试资料为依据,选出最有利于识别地类的最佳波段进行组合分析;另一种方法是计算图像的信息量大小来选定。

但要计算各波段图像的信息量是比较麻烦的,为避免信息量的计算,采用美国查维茨提出的最佳指数(OIF)的概念,即

$$OIF = \sum_{i=1}^3 S_i / \sum_{j=1}^3 |R_{ij}| \quad (2)$$

式中  $S_i$  为第  $i$  个波段的标准差。因为标准差越大,表明数据的离散程度越大,所包含的信息量也越大,而波段相关系数  $R_{ij}$  越小,表明图像数据的独立性越高,信息的冗余度越小。因此,OIF 越大,则该组合图像的信息量越大,说明该组合方案是最优的<sup>[1]</sup>。

#### (1) TM 波段图像的假彩色合成

根据波段选择函数  $P(H,R)$  的计算,少林区最佳的波段组合是 TM4,7,3;多林区则是 TM4,7,2。但由于各波段亮度值分布范围窄,又集中在低亮度区,因此合成图像偏暗,难以作判读分析,必须对各波段作适当的增强处理才行。

#### (2) KL 变换、KT 变换特征图像的合成

前面已对 KL 变换、KT 变换的合成图像作了分析,从图像总体信息量和判读效果考虑,前 3 个组分图像的合成,其信息量最大,可判性能最好。但对某些地类而言则不一定是这样。如多林区比值图像子集经 KL 变换后的组分图像中,第一、第二、第六组分图像的合成,其乔、灌、草之间的区别更明显,植被类型间的判读性能更好;少林区比值图像的 KL 变换中,第一、第二、第五组分图像的合成,突出了水体的特性,可分辨出水的深浅变化。可见,对于专题信息的提取,应选用专题信息集中的组分图像进行合成,以提高目视判读的效果。

#### (3) 比值图像的合成

在对 TM 图像作大量的比值处理的基础上,根据比值图像的直方图形态选用适当的增强方法,使得亮度值分布在 0—255 区间。然后综合分析标准差大小、直方图分布及形态、图像的色调差异、地类的可判性等,选取较好的比值图像进行统计分析,由(2)式计算组合图像的最佳指数 OIF,选择比值合成方案。图版 I 图 5 是多林区比值图像 4/3, (5×7)/3, (3×4)/2 的合成。图像中消除了地形阴影的影响,突出了乔、灌、草之间的差异及水系结构,判读效果较好。

#### (4) 混合图像的合成

不同处理方法、不同组合方法的图像上,可区分的地类及判读的难易程度不同。混合图像是在分析各波段对地类的可判性、标准差、直方图等的基础上,对作为混合图像的子集进行统计分析,并根据(2)式计算的 OIF 指数。选择混合图像的最佳组合方案。如在西吉县的(4×5)/3, (5×7)/4, TM3 拉伸的合成图像,多林的六盘山应用 KT, PC2, (5×7)/3; TM4 拉伸, 5/4, (3×4)/2 的合成图像等。由于混合图像起到了信息特征互补的作用,图像色调层次丰富,整体判读效果好(图版 I 图 6)。

## 参 考 文 献

- [1] 朱启疆等,遥感图像中森林植被信息及其彩色显示优化子集选取方法试验研究,科学出版社,1988。
- [2] J. A. Richards, An Introduction: Remote Sensing Digital Image Analysis, Springer-Verlag, 1986。
- [3] R. A. Schowengerdt, Techniques for Image processing and Classification in Remote Sensing, New York, 1983。

- [4] J. C. Price, Calibration of satellite radiometers and the comparison of vegetation indices, *Remote Sensing of Environment*, 21: 15—27, 1987.

## A STUDY ON THE ANALYSIS OF TM IMAGE INFORMATION CONTENT AND THE EXTRACTION OF CHARACTERISTIC INFORMATION

Lu Dengsheng You Xianxiang Cui Saihua

(*Beijing Forestry University*)

### Abstract

Analysis of image information content is the basis of image treatment. The image information content of three areas of different vegetation cover types namely multi-forest region (more than 40% of forest cover), hilly region of average forest land distribution (about 10—30% of forest cover), and hilly and plain region mainly with farmland was studied, the image of the same region in winter and summer were analyzed. On the basis of image information content analysis, methods of treatment of different images were studied, and their practicality was analyzed and explored as well.

The analysis of information feature shows the general law that the information content of infrared band is higher than that of visible light band, in which the maximum of information content is band TM5, and the minimum, band TM2. Meanwhile, in different conditions, the statistical feature value of correlation between bands, mean value and standard deviation was analyzed, thus making image treatment establish on the basis of the analysis of information content and statistical feature. Accordingly, the enhancement of image, the method of information feature extraction (such as the analysis of principal composition, KT transformation and ratio method) and the combination of bands were researched in a systematic way, and the conditions of their practicality were also explored and evaluated.

**Key words** analysis of information content analysis of statistical feature value enhancement of image information feature extraction combination of bands

图5 少林区(西吉)经比值、KL变换后的组合图像 图中白色为水体,深蓝色为裸露地,浅黄色为草地、蓝绿色、紫红色为农田、饱和的黄色为灌木林,红色为杨树林地

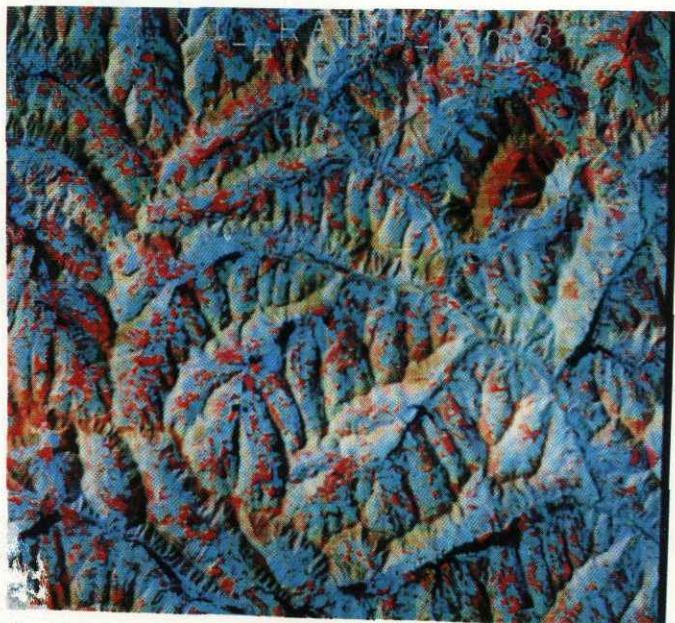
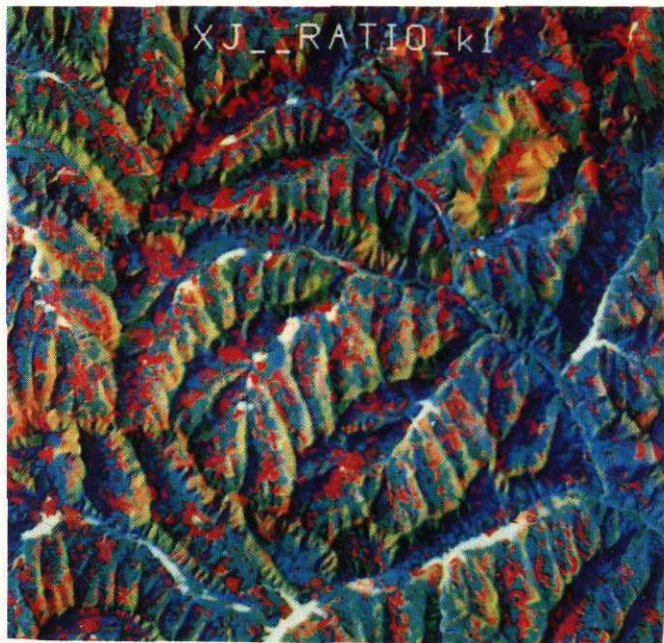


图6 同图5地区经

$$\frac{4 \times 5}{3} \quad \frac{5 \times 7}{4}$$

TM<sub>3</sub> 拉伸后的合成图像黑色为水体红色为林地,褐红色为灌木林、白色为裸露地,浅蓝色为农田