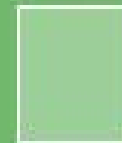


测绘遥感信息工程系



遥感图像分类处理



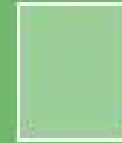
遥感影像判读



遥感影像分类



分类概述

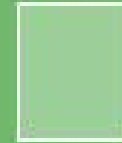


- 分类的定义

- 遥感图像分类以**遥感数字图像**为研究对象，在计算机系统支持下，综合运用地学分析、遥感图像处理、地理信息系统、模式识别与人工智能技术，实现地学专题信息的智能化获取。
- 其基本目标是将人工目视解译遥感图像发展为计算机支持下的遥感图像理解。



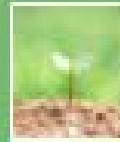
分类概述



- 分类的意义
- 由于利用遥感图像可以客观、真实和快速地获取地球表层信息，这些现势性很强的遥感数据在自然资源调查与评价、环境监测、自然灾害评估与军事侦察上具有广泛应用前景。因此，利用计算机进行遥感图像智能化解译，快速获取地表不同专题信息，并利用这些专题信息迅速地更新地理数据库，这是实现遥感图像自动理解的基础研究之一，因此具有重要的理论意义和应用前景。



分类概述

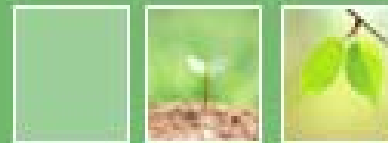


● 分类的难点

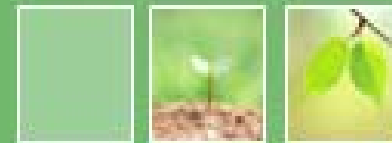
利用计算机对遥感数字图像进行分类难度很大。

- ✓ 遥感图像是从遥远的高空成像的，成像过程要受传感器、大气条件、太阳位置等多种因素的影响。影像中所提供的目标地物信息不仅不完全，而且或多或少地带有噪声，因此人们需要从不完全的信息中尽可能精确地提取出地表场景中感兴趣的目标物。
- ✓ 遥感影像信息量丰富，与一般的图像相比，其包容的内容远比普通的图像多，因而内容非常“拥挤”。不同地物间信息的相互影响与干扰使得要提取出感兴趣的目标变得非常困难。
- ✓ 遥感图像的地域性、季节性和不同成像方式更增加了计算机对遥感数字图像进行解译的难度。

图像分类的原理



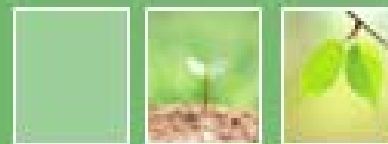
1. 计算机遥感图像分类是**统计模式识别技术**在遥感领域中的具体应用。统计模式识别的关键是提取待识别模式的一组统计特征值，然后按照一定准则作出决策，从而对遥感图像予以识别。
2. 遥感图像分类的主要依据是**地物的光谱特征**，即地物电磁波辐射的测量值，这些测量值可以用作遥感图像分类的原始特征变量。然而，就某些特定地物的分类而言，多波段影像的原始亮度值并不能很好地表达类别特征，因此需要对数字图像进行运算处理（如比值处理、差值处理、主成分变换以及K-T变换等），以寻找能有效描述地物类别特征的模式变量，然后利用这些特征变量对数字图像进行分类。分类是对图像上每个像素按照亮度接近程度给出对应类别，以达到大致区分遥感图像中多种地物的目的。



3. 遥感图像分类的常使用**距离**和**相关系数**来衡量相似度。

- 采用距离衡量相似度时，距离越小，相似度越大。
- 采用相关系数衡量相似度时，相关程度越大，相似度越大。

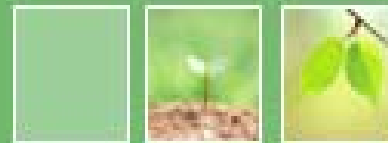
遥感图像分类基本过程



遥感数字图像计算机分类基本过程如下：

- 首先明确遥感图像分类的目的及其需要解决的问题，在此基础上根据应用目的选取特定区域的遥感数字图像，图像选取时应考虑图像的空间分辨率、光谱分辨率、成像时间、图像质量等。
- 根据研究区域，收集与分析地面参考信息与有关数据。为提高计算机分类的精度，需要对数字图像进行辐射校正和几何纠正（这部分工作也可能由提供数字图像的卫星地面站完成）。
- 对图像分类方法进行比较研究，掌握各种分类方法的优缺点，然后根据分类要求和图像数据的特征，选择合适的图像分类方法和算法。根据应用目的及图像数据的特征制定分类系统，确定分类类别，也可通过监督分类方法，从训练数据中提取图像数据特征，在分类过程中确定分类类别。

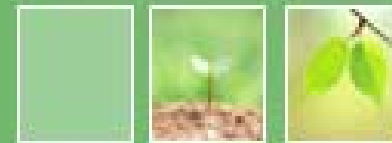
遥感图像分类基本过程



- 找出代表这些类别的统计特征。
- 为了测定总体特征，在监督分类中可选择具有代表性的训练场地进行采样，测定其特征。在无监督分类中，可用聚类等方法对特征相似的像素进行归类，测定其特征。
- 对遥感图像中各像素进行分类。包括对每个像素进行分类和对预先分割均匀的区域进行分类。
- 分类精度检查。在监督分类中把已知的训练数据及分类类别与分类结果进行比较，确认分类的精度及可靠性。在非监督分类中，采用随机抽样方法，分类效果的好坏需经实际检验或利用分类区域的调查材料、专题图进行核查。
- 对判别分析的结果统计检验。



遥感图像分类方法



利用遥感图像进行分类（classification）是以区别图像中所含的多个目标物为目的的，对每个像元或比较匀质的像元组给出对应其特征的名称。在分类中注重的是各像元的灰度及纹理等特征。分类方法主要包括以下三大类：

- 非监督分类法
- 监督分类法
- 新的探索：模糊分类法、面向对象分类法等。



非监督分类

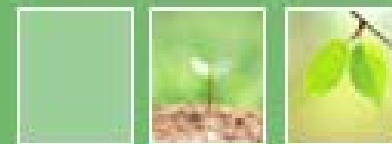
非监督分类方法是在没有先验类别（训练场地）作为样本的条件下，即事先不知道类别特征，主要根据统计性判别准则，以像元间相似度的大小进行归类合并（即相似度的像元归为一类）的方法。它的目的是使得属于同一类别的像素之间的距离尽可能的小而不同类别上的像素间的距离尽可能的大。

其常用方法有：

- C均值法
- ISODATA算法



非监督分类



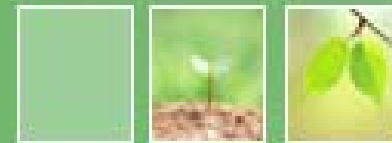
C均值法

C均值算法将像素划分到光谱距离最小的类。分类过程开始于特定数量的随机聚类中心或已存在的模版均值。分类过程不断重复，直至聚类中心接近于数据中各个类的中心为止。 算法具体流程如下：

- ① 任选c个模式特征矢量作为初始聚类中心： $z_1^{(0)}, z_2^{(0)}, \dots, z_c^{(0)}$
- ② 将待分类的模式特征矢量集 $\{x_i\}$ 中的模式逐个按最小距离原则分给c类中的某一类，即 如果， $d_i^{(k)} = \min_j [d_{ij}^{(k)}]$ $i=1, 2, \dots, N$ 则判 $x_i \in \omega_i^{(k+1)}$
- ③ 计算重新分类后的各类心 $z_j^{k+1} = \frac{1}{N_j^{(k+1)}} \sum_{x_i \in \omega_j^{(k+1)}} x_i \quad j=1, 2, \dots, c$
- ④ 如果 $z_j^{(k+1)} = z_j^{(k)}$ ($j=1, 2, \dots, c$)，则结束；否则， $k=k+1$ ，转至 (2)



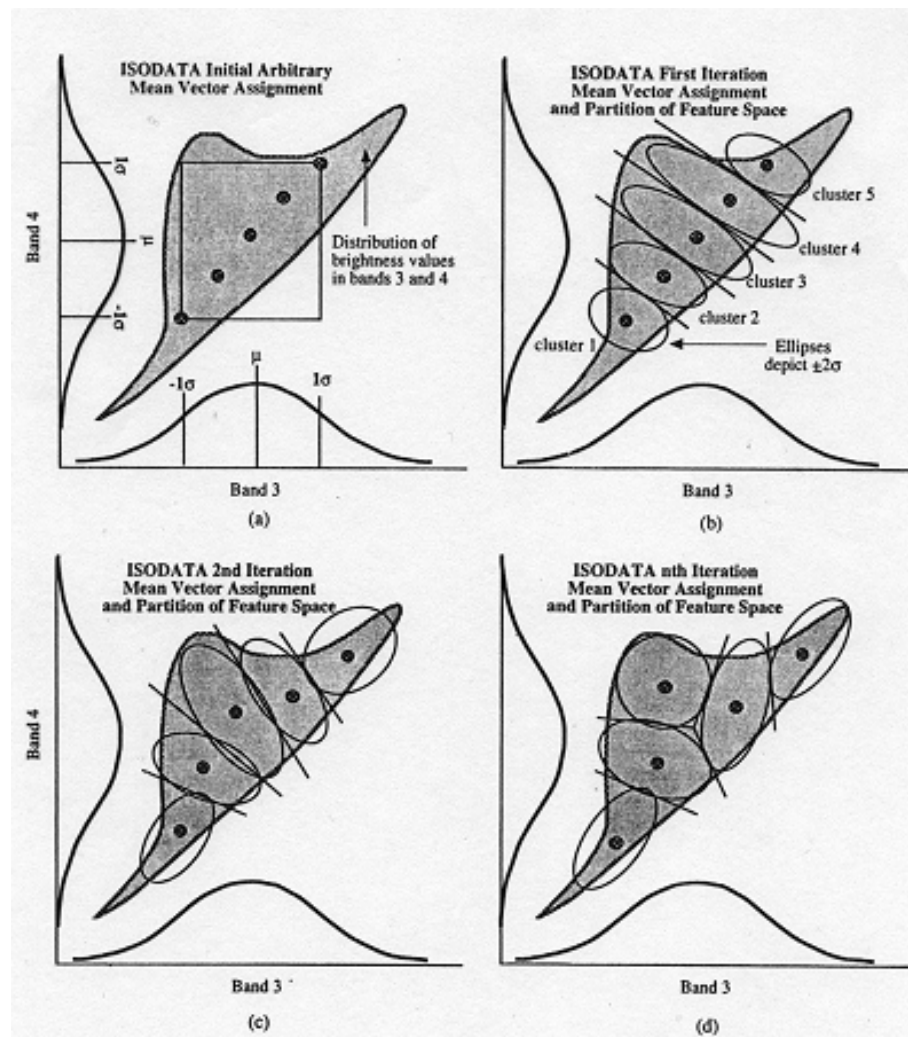
非监督分类



- ISODATA算法是自组织方式的，因为它几乎不需要人工输入。成熟的ISODATA算法通常需要分析人员确定如下标准：
- C_{max} ：算法确定的最大类别数；
- T ：两次迭代之间，类值不变像元所占的最大百分比；
- M ：ISODATA迭代运算的最大次数。当达到该值时，算法终止；
- 每类中最少像元数：如果某类像元占总像元数的百分比小于最小容许值，就删除该类，并把属于它的像元分配到另一类中；
- 最大标准差：当某类的标准差超过制定的最大标准差，且类中的像元数大于制定的最少像元数的2倍时，该类就分解为2个；
- 最大合并类数：进行类合并过程中，每次最多可以合并的类别数目；
- 类间最小距离：如果2个类之间的加权距离小于该值，就将这2个类合并。

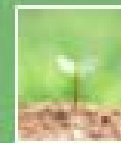
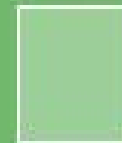
非监督分类

- 动态聚类就是在开始时先建立一批初始中心，而让待分类的各个像元依据某些判决准则向初始中心凝聚，然后再逐步修改调整中心，重新分类；并根据各类离散性统计量和两类间可分离性的统计量再进行合并和分裂。此后再修改调整中心，这样不断继续下去，直到分类比较合理为止。





非监督分类



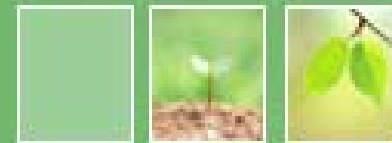
- 在用计算机编制分类程序时，初始聚类中心可按如下方式确定：设初始类别数为 n ，这样共有 n 个初始聚类中心，求出图像的均值 M 和方差 σ ，按下式可求出初始聚类中心：

$$\bar{x}_k = M + \sigma \left(\frac{2(k-1)}{n-1} - 1 \right)$$

- $k=1, 2, \dots, n$ ，为初始类中心编号， n 为初始类总数。

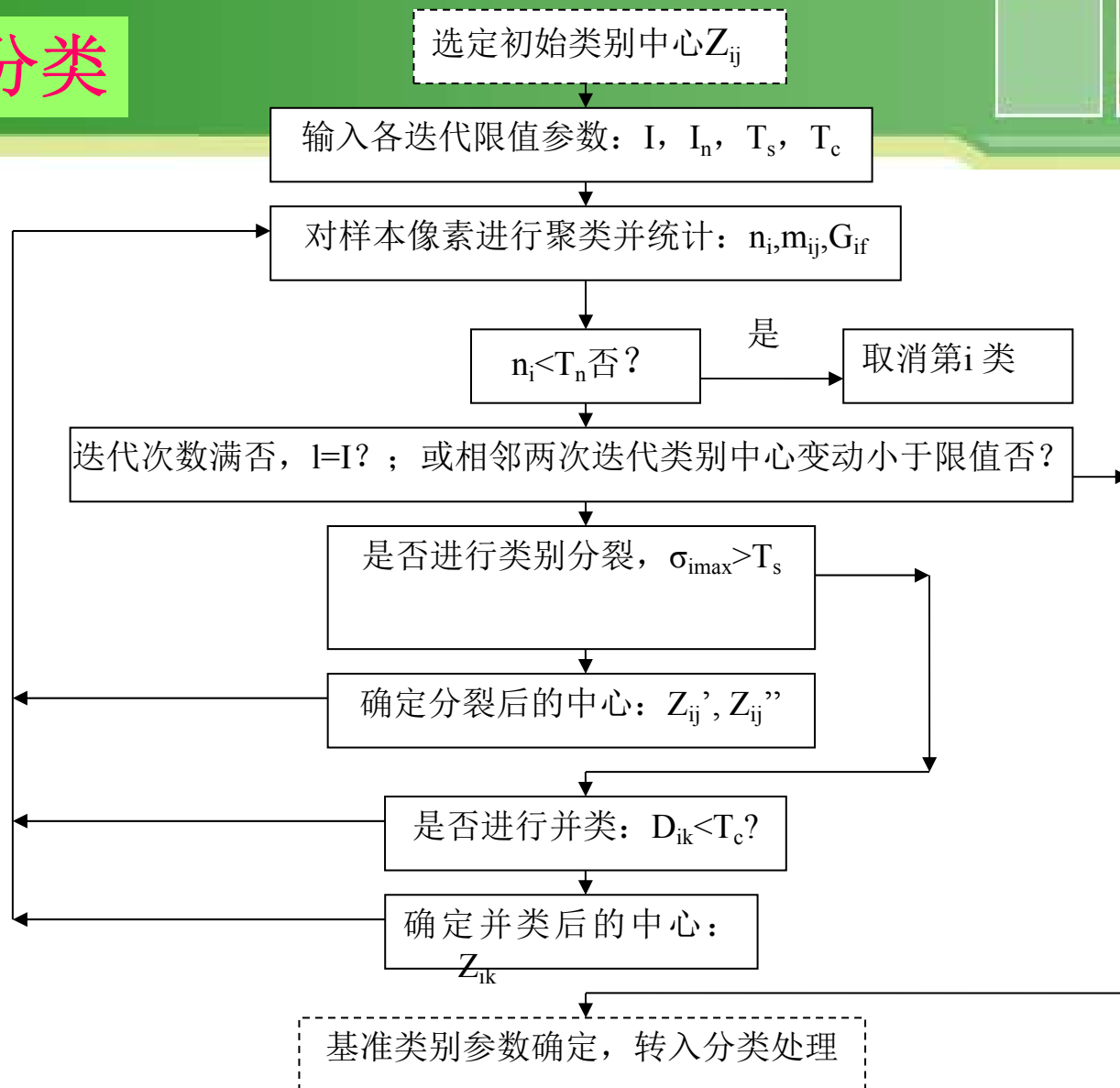


非监督分类



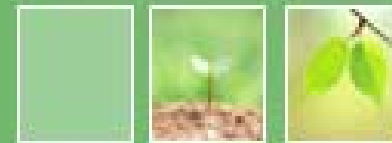
- 计算像素与初始类别中心的距离，把该像素分配到最近的类别中。动态聚类法中类别间合并或分割所使用的判别标准是距离，待分像元在特征空间中的距离说明互相之间的相似程度，距离越小，相似性大，则它们可能会归入同一类。
- 计算并改正重新组合的类别中心，如果重新组合的像素数目在最小允许值以下，则将该类别取消，并使总类别数减1。当类别数在一定的范围，类别中心间的距离在阈值以上，类别内的方差的最大值为阈值以下时，可以看做动态聚类的结束。当不满足动态聚类的结束条件时，就要通过类别的合并及分离，调整类别的数目和中心间的距离等，然后返回到上一步，重复进行组合的过程。

非监督分类



ISODATA法过程框图

非监督分类



- 动态聚类法中有类别的合并或分裂，这说明迭代过程中类别总数是可变的。
- 如果两个类别的中心点距离近，说明相似程度高，两类就可以合并成一类；或者某类像元数太少，该类就要合并到最相近的类中去。

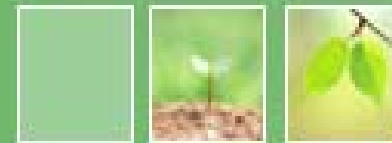
$$D_{ik} = \sum_{j=1}^n |M_{ij} - M_{kj}|$$

- 新类别的均值和方差如下式：

$$M_{(i+k)j} = \frac{n_i M_{ij} + n_k M_{kj}}{n_i + n_k}$$

$$\sigma_{(i+k)j}^2 = \frac{1}{n_i + n_k} \left\{ n_i \left[\sigma_{ij}^2 + (M_{ij} - M_{(i+k)j})^2 \right] + n_k \left[\sigma_{kj}^2 + (M_{kj} - M_{(i+k)j})^2 \right] \right\}$$

非监督分类



- 类别的分裂也有两种情况：
 - 某一类像元数太多，就设法分成两类；
 - 如果类别总数太少，就将离散性最大的一类分成两个类别，可以先求出每个类别的均值和标准差，然后通过每一个波段的标准偏差设定阈值来实现，标准差大于阈值，该类就要分裂。
 - 重新定义新的两类的均值 M'_{ij} 、 M''_{ij} 和方差 σ'_{ij} 、 σ''_{ij} ：

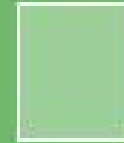
$$M'_{ij} = M_{ij} + \frac{1}{3}\sigma_{ij}$$

$$M''_{ij} = M_{ij} - \frac{1}{3}\sigma_{ij}$$

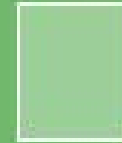
$$\sigma'_{ij} = \sigma''_{ij} = \frac{2}{3}\sigma_{ij}$$



监督分类



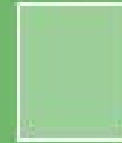
- 选择具有代表已知地面覆盖类型的**训练样本区**，用训练样本区中已知地面各类地物样本的光谱特性来“训练”计算机，获得识别各类地物的判别函数或模式（如均值、方差、判别域等），并以此对未知地区的像元进行分类处理，分别归入到已知具有最大相似度的类别中。
- **最小距离分类法、最近邻分类法、多级切割分类法、最大似然比分类法等。**



- 最小距离分类法
 - 最小距离分类法概念
 - 最小距离分类法判据
 - 最小距离分类法特点
 - 最小距离分类法示例

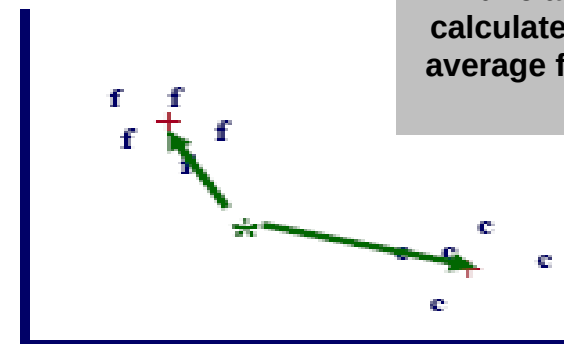
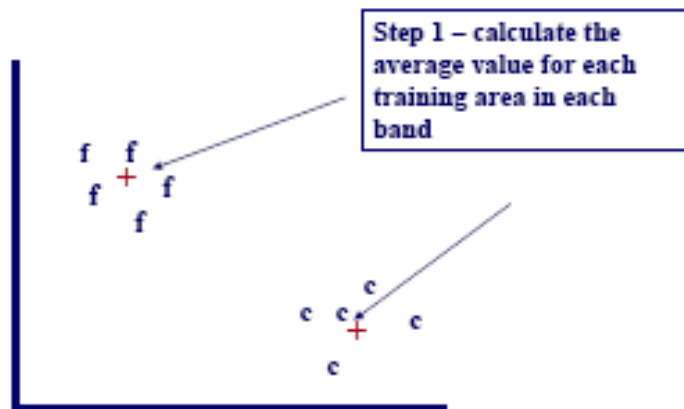


监督分类



• 最小距离分类法

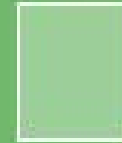
- 是用特征空间中的距离表示像元数据和分类类别特征的相似程度，在距离最小时（相似度最大）的类别上对像元数据进行分类的方法。



* - Unclassified pixel

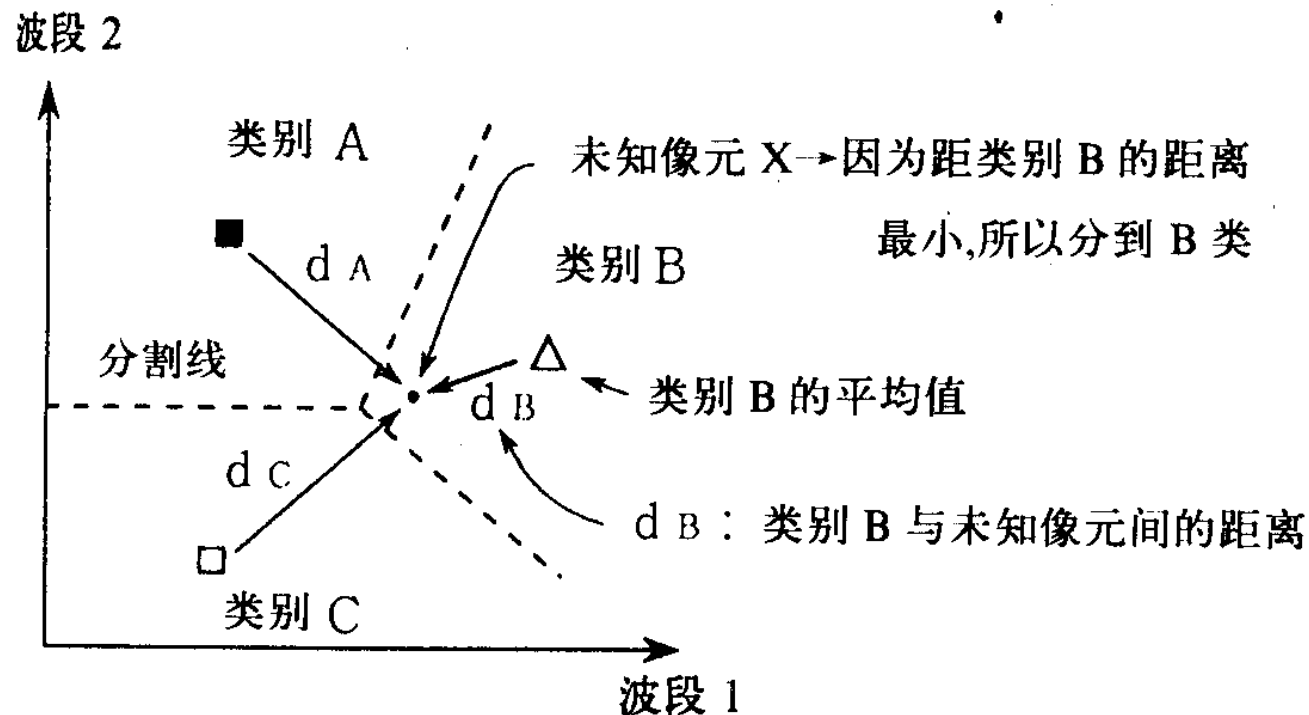


监督分类



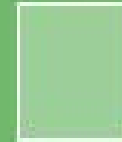
• 最小距离分类法判据

- 这种方法要求对遥感图像中每一个类别选一个具有代表性的统计特征量（均值），首先计算待分像元与已知类别之间的距离，然后将其归属于距离最小的一类。

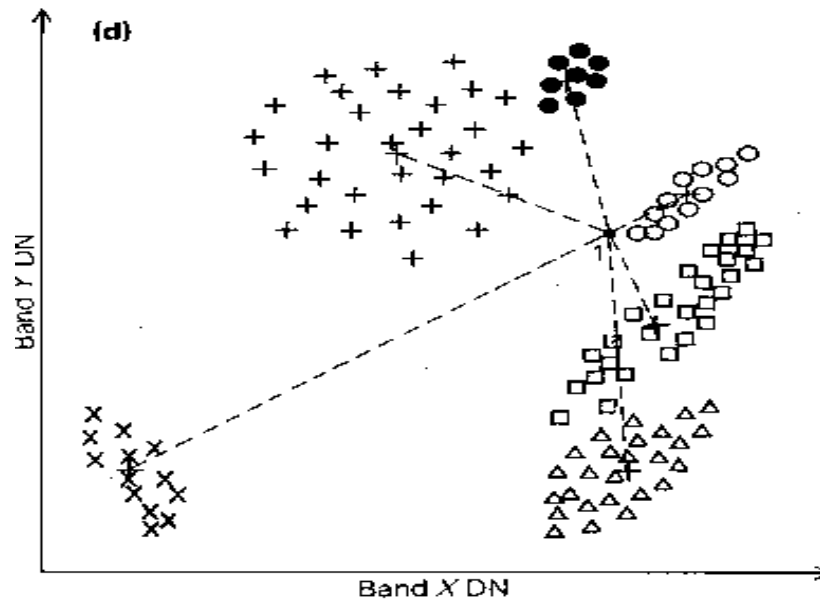




监督分类



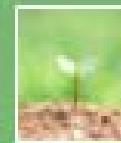
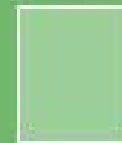
• 最小距离分类法-示例



	<i>Key</i>
	corn
	forest
	hay
	sand
	urban
	water



监督分类

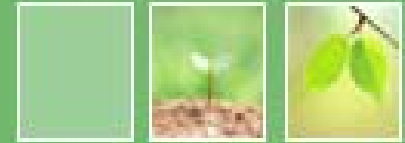


- 最小距离分类法-特点

- 优点：最小距离分类法原理简单，计算速度快；
- 缺点：分类精度不高，没有解决分类的边界问题；
- 适用对象：它可以在快速浏览分类概况中使用；



监督分类

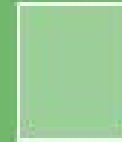


- 最大似然比分类法

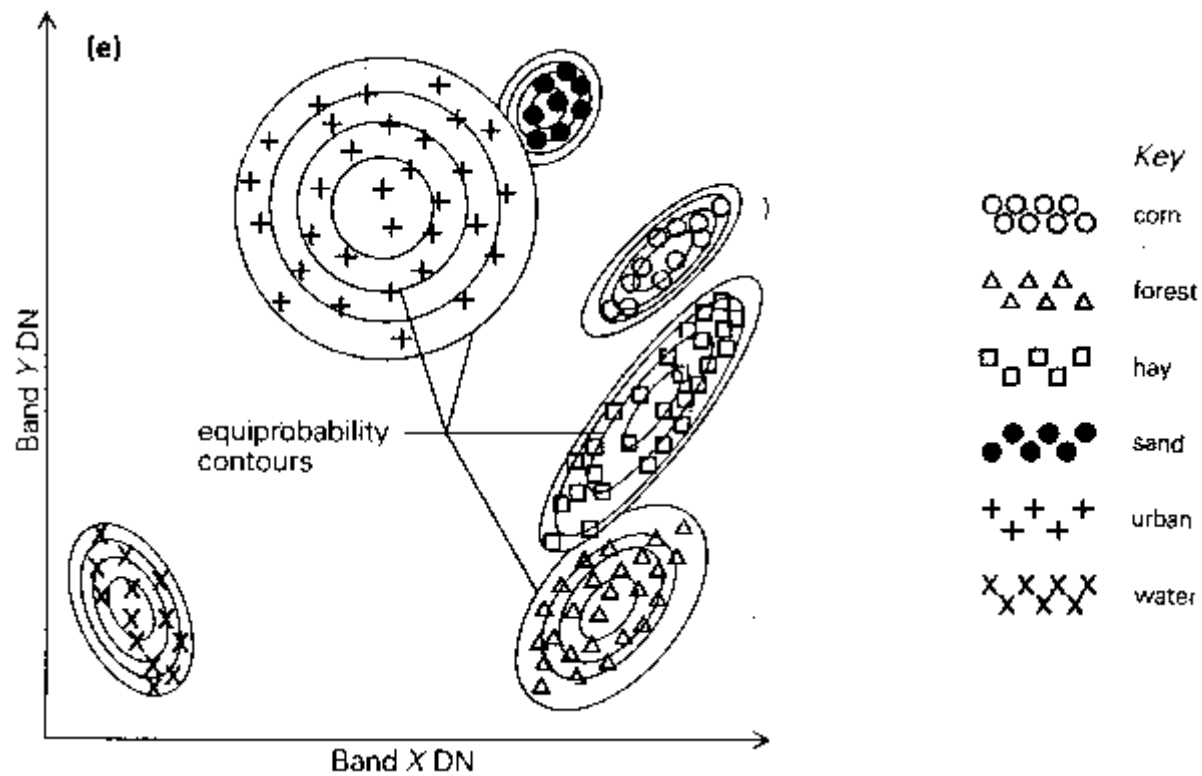
- 最大似然比分类法 (maximum likelihood classifier) 是应用非常广泛的监督分类之一。
- 主要是利用概率密度函数，求出每个像素对于各类别的似然度 (likelihood)，把该像元分到似然度最大的类别中去的方法。
- 它假定训练区地物的光谱特征和自然界大部分随机现象一样，近似服从正态分布，利用训练区可求出均值、方差以及协方差等特征参数，从而可求出总体的先验概率密度函数。当总体分布不符合正态分布时，其分类可靠性将下降，这种情况下不宜采用最大似然比分类法。

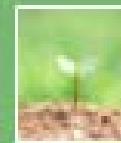
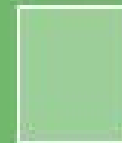


监督分类



- 最大似然比分类法





- 最大似然比分类法

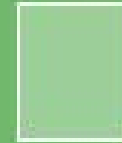
最大似然比分类法在多类别分类时，常采用统计学方法建立起一个判别函数集，然后根据这个判别函数集计算各待分像元的归属概率。

$$L_k = P(k | x) = P(k) \times P(x | k) / \sum P(i) \times P(x | i)$$

x 为待分像元， $P(k)$ 为类别 k 的先验概率，可以通过训练区来决定。由于上式中分母和类别无关，在类别间比较的时候可以忽略。



监督分类



- 监督/非监督分类方法比较

- 相比之下，非监督分类不需要更多的先验知识，它根据地物的光谱统计特性进行分类。因此，非监督分类方法简单，且分类具有一定的精度。
- 严格说来，分类效果的好坏需要经过实际调查来检验。当光谱特征类能够和唯一的地物类型（通常指水体、不同植被类型、土地利用类型、土壤类型等）相对应时，非监督分类可取得较好分类效果。当两个地物类型对应的光谱特征类差异很小时，非监督分类效果不如监督分类效果好。

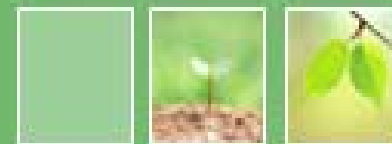


分类后处理

无论是监督分类还是非监督分类，其结果都会产生一些面积很小的图斑。无论从专题制图的角度还是实际应用的角度，都有必要对这些小图斑进行剔除。处理方法：

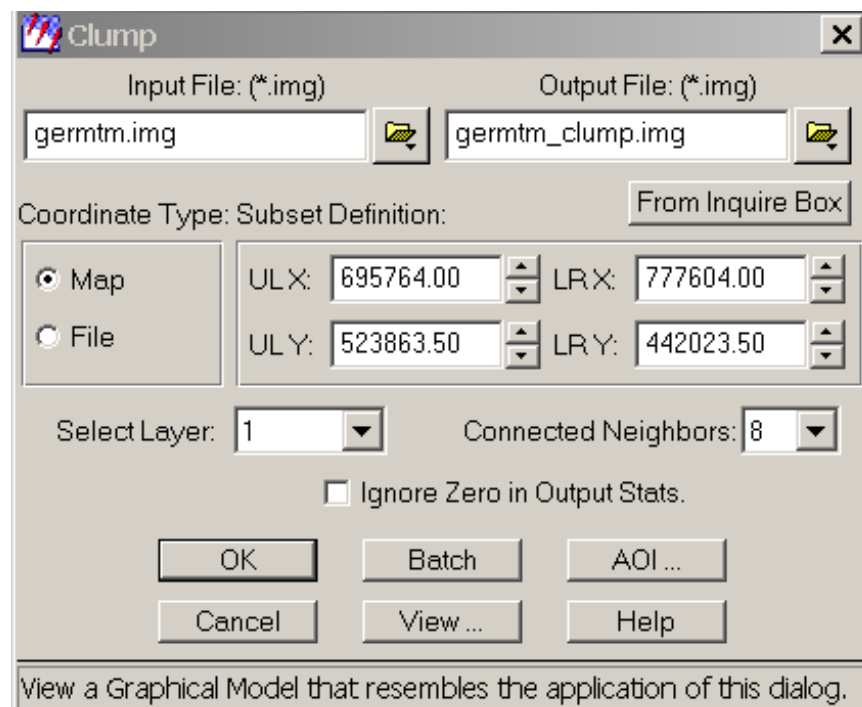
- 聚类统计(Clump)
- 过滤分析(Sieve)
- 去除分析(Eliminate)
- 分类重编码 (Recode)

分类后处理

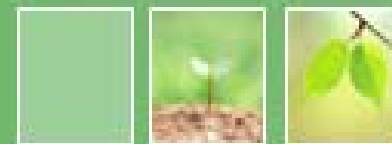


- 聚类统计

- 通过对分类专题图像计算每个分类图斑的面积、记录相邻区域中最大图斑面积的分类值等操作，产生一个Clump类组输出图像，其中每个图斑都包含Clump类组属性。这是一个中间结果，供下一步处理使用。

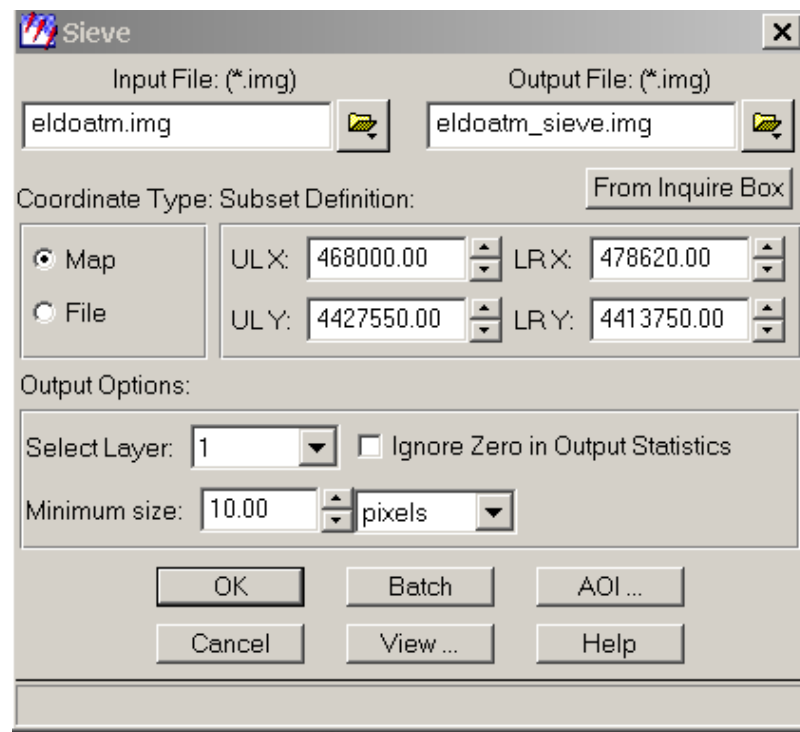


分类后处理

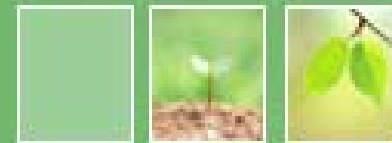


- 过滤分析

- 对经C1ump处理后的C1ump类组图像进行处理，按照定义的数值大小，删除C1ump图像中较小的类组图斑，并给所有小图斑赋予新的属性值0。显然，这引出了一个小图斑归属问题。可以与原分类图对比确定新属性。

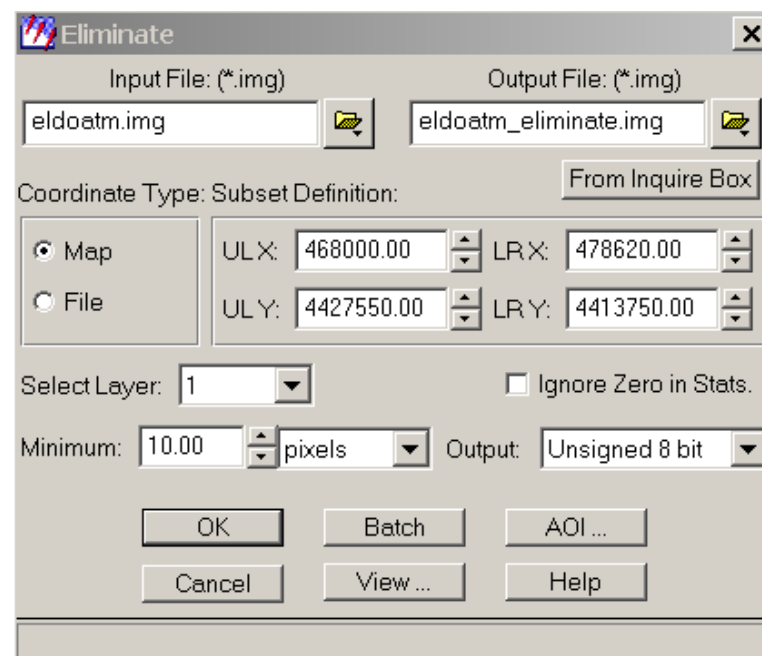


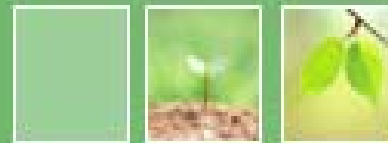
分类后处理



- 去除分析

- 用于删除原始分类图像中的小图斑或小Clump类组，与过滤不同，去除将删除的小图斑合并到相邻的最大分类中，而且如果输入图像是Clump聚类图像，经过去除处理后，将分类图斑的属性值自动恢复为Clump处理前的原始分类编码。即结果是简化的分类图像。





- 分类重编码

- **分类重编码：**主要是针对非监督分类而言的，因在非监督分类过程中，用户一般要定义比最终需要多一定数量的分类数；在完全按照像元灰度值通过ISODATA聚类获得分类方案后，首先是将专题分类图像与原始图像对照，判断每个类别的专题属性，然后对相似或类似的分类通过图像重编码进行合并，并定义分类名称和颜色。
- 分类重编码还可以用在其它方面，作用有所不同。

分类后处理

- 分类重编码

The image shows two overlapping dialog boxes from a GIS application. The 'Recode' dialog is in the background, and the 'Thematic Recode' dialog is in the foreground.

Recode Dialog:

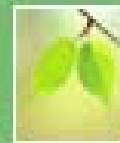
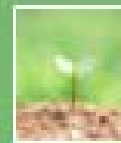
- Input File: (*.img)
- Output File: (*.img)
- Setup Recode ...
- ☐ Ignore Zero in Stats.
- Input: Unsigned 4
- Output: Unsigned 4
- Buttons: OK, Cancel, Batch, View ..., AOI ..., Help

Thematic Recode Dialog:

Value	New Value	Red	Green	Blue	Class_Names
0	0	0.000	0.000	0.000	
1	1	1.000	0.784	0.000	Riparian
2	1	0.588	1.000	0.588	Grassland and Scrub
3	2	0.196	0.588	0.196	Chaparral
4	2	0.196	1.000	1.000	Wetlands
5	3	0.392	0.863	0.784	Emergent Vegetation
6	3	0.392	0.529	0.784	Water

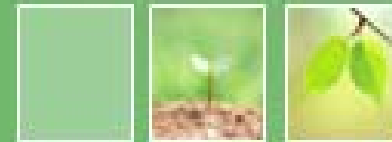
New Value: Change Selected Rows

Buttons: OK, Cancel, Help



- 遥感图像分类精度检查

- 在监督分类中把已知的训练数据及分类类别与分类结果进行比较，确认分类的精度及可靠性。
- 在非监督分类中，采用随机抽样方法，分类效果的好坏需经实际检验或利用分类区域的调查材料、专题图进行核查。
- 误差矩阵的评价



• 遥感图像分类精度检查-误差矩阵

实际类别	试验像元的百分率 (%)			试验像元	
	类别1	类别2	类别3		
1	84.3	4.9	10.8	100%	102
2	8.5	80.3	11.2	100%	152
3	6.1	4.1	89.8	100%	49

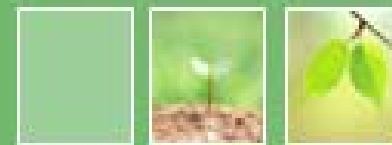
对角线元素便是正确分类，非对角线便是错误分类；

根据误差矩阵可算出平均精度： $S = (84.3\% + 80.3\% + 89.8\%) / 3 = 84.8\%$

由于各类别样本总数不一样，所以往往使用加权平均：

$$S = (84.3\% * 102 + 80.3\% * 152 + 89.8\% * 49) / (102 + 152 + 49) = 83.2\%$$

遥感图像分类中存在的问题

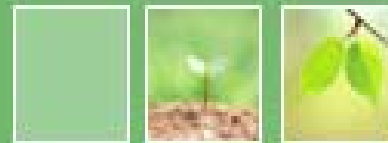


遥感图像分类算法设计的主要依据是地物光谱数据。因此，存在着如下的问题：

未充分利用遥感图像提供的多种信息

- 遥感数字图像计算机分类的依据是像素具有的多光谱特征，并没有考虑相邻像素间的关系。例如，被湖泊包围的岛屿，通过分类仅能将陆地与水体区别，但不能将岛屿与临近的陆地（假定二者地面覆盖类型相同，具有同样的光谱特征）识别出来。
- 这种方法的主要缺陷在于地物识别与分类中没有利用到地物空间关系等方面的信息。

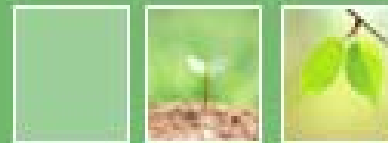
遥感图像分类中存在的问题



➤ 未充分利用遥感图像提供的多种信息

- 统计模式识别以像素作为识别的基本单元，未能利用图像中提供的形状和空间位置特征，其本质是地物光谱特征的分类。例如，根据水体的光谱特征，在分类过程中可以识别构成水体的像素，但计算机无法确定一定空间范围的水体究竟是湖泊还是河流。这个问题如果引入地物形状特征则可以识别。
- 显然，遥感图像计算机分类未能充分利用遥感图像提供的多种信息。因此图像分类后，可以利用分类的结果，将这些目标对象进行重组，在区域分割或边界跟踪的基础上抽取遥感图像形态、纹理特征和空间关系等特征，然后利用这些特征对图像进行解译。

遥感图像分类中存在的问题



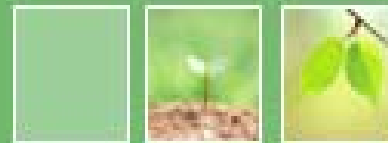
➤ 提高遥感图像分类精度受到限制

这里的分类精度是指分类结果的正确率。正确率包括地物属性被正确识别，以及它们在空间分布的面积被准确度量。遥感数字图像分类结果在没有经过专家检验和多次纠正的情况下，分类精度一般不超过90%，其原因除了与选用的分类方法有关外，还存在着制约遥感图像分类精度的几个客观因素：

- 大气状况的影响

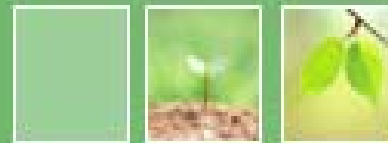
不少人理想化地认为遥感图像只记录遥感观测区域内的地物电磁辐射能量，遥感图像的灰度大小及其变化只反映了地物的辐射光谱特征变化，这种观念是不正确的。

遥感图像分类中存在的问题



- 地物辐射电磁波，必须经过大气层才能到达传感器，大气的吸收和散射会对目标地物的电磁波产生影响，其中大气吸收使得目标地物的电磁波辐射被衰减，到达传感器的能量减少，散射会引起电磁波行进方向的变化，非目标地物发射的电磁波也会因为散射而进入传感器，这样就导致遥感图像灰度级产生一个偏移量。
- 对多时相图像进行分类处理时，由于不同时间大气成分以及湿度不同，散射影响也不同，因此遥感图像中的灰度值不完全反映目标地物辐射电磁波的特征。为了提高遥感图像分类的精度，必须在图像分类以前进行大气纠正。

遥感图像分类中存在的问题

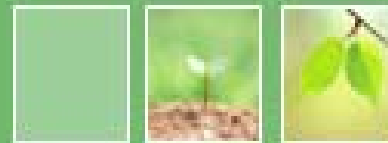


- 下垫面的影响

下垫面的覆盖类型和起伏状态对分类具有一定影响。下垫面的覆盖类型多种多样，受传感器空间分辨率限制，农田中的植被、土壤和水渠，石质山地稀疏的灌丛和裸露的岩石均可以形成混合像元，它们对遥感图像分类的精度影响很大。

这种情况可以在分类前首先进行混合像元分解，把它们分解成子像元后再分类。分布在山区向阳面与背阳面的同一类地物，单位面积上接收太阳光能不同，地物电磁波辐射能量也不同，其灰度值也存在差异，容易造成分类错误。在地形起伏变化较大时，可以采用比值图像代替原图像进行分类，以消除地形起伏的影响。

遥感图像分类中存在的问题



- 其他因素的影响

- ✓ 图像中的云朵会遮盖目标地物的电磁波辐射，影响图像分类。对于图像中仅有少量云朵时，分类前可以采用去噪音方法进行清除；
- ✓ 多时相图像分类时，不同景的图像由于成像时光照条件的差别，同一地物电磁波辐射量存在差别，这也会对分类产生影响
- ✓ 地物边界的多样性，使得判定类别的边界往往是很困难的事。例如，湖泊和陆地具有明确的界线，但森林和草地的界线则不明显，不少地物类型间还存在着过渡地带，要精确将其边界区别出来，并非是一件容易的事。

提高遥感图像分类精度，既需要对图像进行分类前处理，也需要选择合适的分类方法。



Thank You!

