

ADS - Project.1

任务：

1. 提取图片的特征值，构成一个多维的矩形，用R树来写出索引
2. 写一个搜索界面，来查询某张图片，给出根据相关性排序的图片列表

Key Details

1. 图片资源存储在data\Image.zip中
2. 每张图片的名称下划线之前的是类别，但是查询的过程中不能使用这个信息
3. 提取特征值：
 - Loading Image: 在目录 src>LoadingImage 里有代码，作用是把图片信息载入主存，然后可以处理每个像素的RGB值
 - 提取特征值算法：Colour Histogram Features (Recommended)
 - 助教提供了一个根据Colour Moment Features 已经提取好的9维特征值，我们可以根据这个来写起步，但是不能用于最后的作业
4. R-树：有一个现成的可以在它的基础上改进即可，代码在src\rtree里，它只能在Unix/Linux或者基于它们的系统上编译，在该目录下运行“make linear”命令就可以生成这个项目。另外这个代码是助教改过的，可以直接用刚提到的9维特征值作为数据集。至于运行的一些参数细节还是看原文档吧QAQ
5. 这个项目允许使用一些R-树的现成的包，助教也提供了两个，在src\rtree_alternative_package目录下
6. 最终效果的评价是根据磁盘访问次数，而不是CPU时间。由于R-树的特性，我们可以根据数结点访问次数来估算这个时间复杂度

要解决的具体问题和评分细则：

1. 测试提取不同的特征值数量（e.g. 4, 8, 12, 16, 20）和不同的插入的数据集大小（e.g. 1000, 2000, 3000, 4000, 5000）下的平均效果（60%）
2. 实现我们觉得最有效的对多种图片特征值提取的方法（可以用现成的也可以自己设计），提取出来的结果应该是一个向量，并且与助教提供的特征数据比较查询效果（20%）
3. 尝试研究提取出的特征对结果的相关度的影响，就是说在不同的特征值生成的索引和查询系统下，计算查询图片和结果图片的相关度（20%）
4. 选做任务（每个10%，最多加20%）：
 - 指出在数据插入的时候影响节点分裂的因素并说明原因，最好提出自己的想法来减少结点分裂
 - 指出在数据插入的时候影响MBR重叠的因素并说明原因，最好提出自己的想法来减少MBR重叠
 - 实现有至少一个你认为有用的相似度函数的查询算法，比较在不同的相似度函数下相似查询的表现（相似度函数大概意思就是通过这个函数能输出一个0-1的小数来表示两个东西的相似度）

- 修改R-树代码，实现一个非MBR为形状的R-树（比如圆形），并测试表现效果
- 其它可能的优化R-树查询效果的方法，并用实际数据做证明

关于提交

- 时间：5月23日23:59:59之前
- 实验报告（中英文均可，3页之内，必须用模板并最后转成pdf，一定要说明引用的库或者包）：

1. 系统的outline
2. 分别阐述对三个必做问题的见解
3. 分析与结论
4. （选做）你对索引的新想法，提供具体方法，并且指出分析结果来支撑你的想法
5. 必要的参考文献

最后，

原文档还写了一些提示资料和参考文献，直接看就好。