数字媒体技术基础

Meng Yang

www.smartllv.com

SUN YAT-SEN University

机器智能与先进计算教育部重点实验室

智能视觉语言学习研究组

Course Outline

- 9 数字媒体检索技术
- 9.1 基于标签的媒体检索
- 9.2 基于内容的媒体检索
- 9.3 多媒体内容的索引
  - 9.3.1 静态相关性TF-IDF
  - 9.3.2 动态相关性PageRank
- 9.4 媒体检索反馈技术

目 录

数字媒体检索技术

9.1 基于标签的媒体检索

生活中常见的标签

电影

图片

生活中常见的标签

音乐

图书

标签（Tag）

- 标签通常使用简单的词汇描述或表示视频、图片和音乐等各种不同形式的多媒体资源。
- 用户可以根据标签与多媒体资源内容的匹配来检索到用户当前需求的资源。

标签（Tag）

- 通常标签分为两种：传统分类法，分众分类法。
- 传统分类法：预定义好一系列标签与资源内容相匹配，用户只能根据定义好的标签进行检索。
- 具有严谨具体的特性。

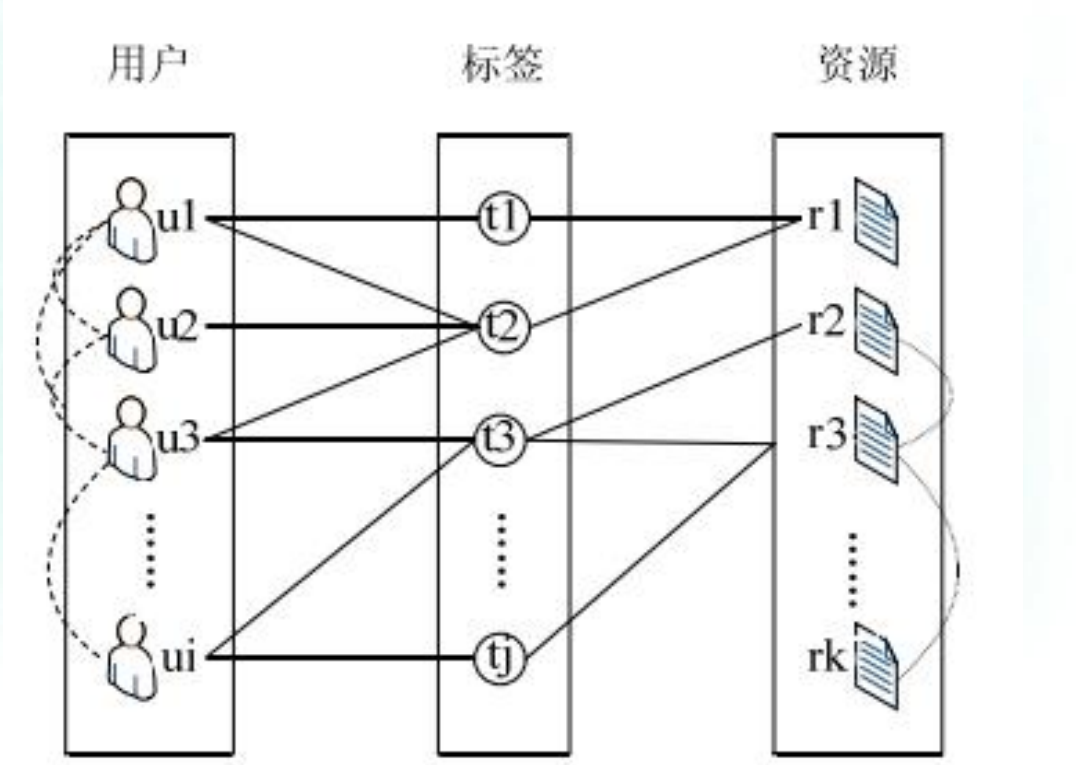
有没有更好的标签方法？

分众分类法（Folksonomy）

- Folksonomy = Folks + Taxonomy
- 基本理念是由所有用户共同选择资源的标签，产生普遍的资源分类标签。
- 具有公开共享、无等级标签分类和动态更新等特性。

分众分类法（Folksonomy）

- 多个用户可以使用一个标签，一个资源可以被多个标签标注。



分众分类法（Folksonomy）

- 根据资源的标签词频，词频越高的标签越能代表资源的内容。

C++ Primer 中文版（第 5 版）

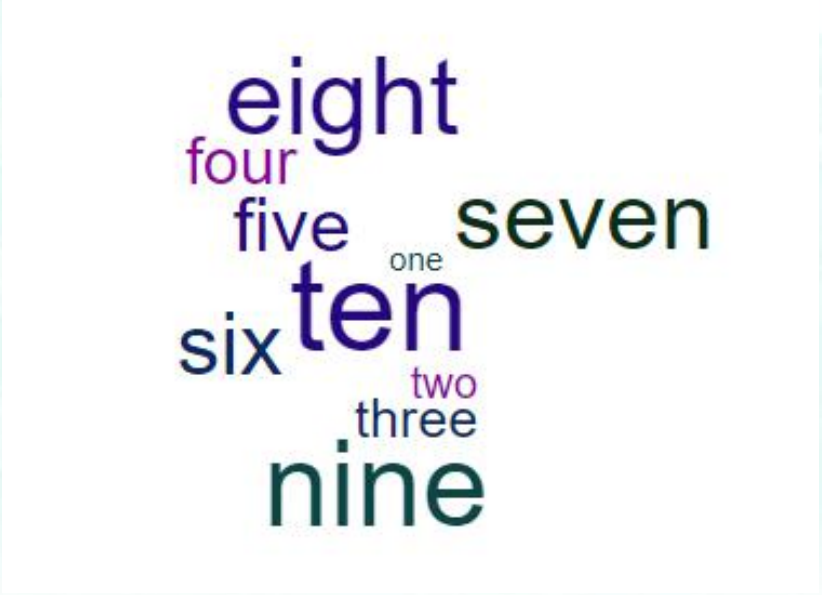
作者: [美] Stanley B. Lippman / [美] Josée Lajoie / [美] Barbara E. Moo
出版社: 电子工业出版社
出品方: 博文视点
原作名: C++ Primer, 5th Edition
译者: 王刚 / 杨巨峰
出版年: 2013-9-1
页数: 838
定价: CNY 128.00
装帧: 平装
ISBN: 9787121155352

豆瓣成员常用的标签(共202个)

C++ 编程 计算机 C++11 C/C++ 编程语言 经典 程序设计

标签云

- 标签词频的可视化，词频越大的标签在标签云中字体越大。



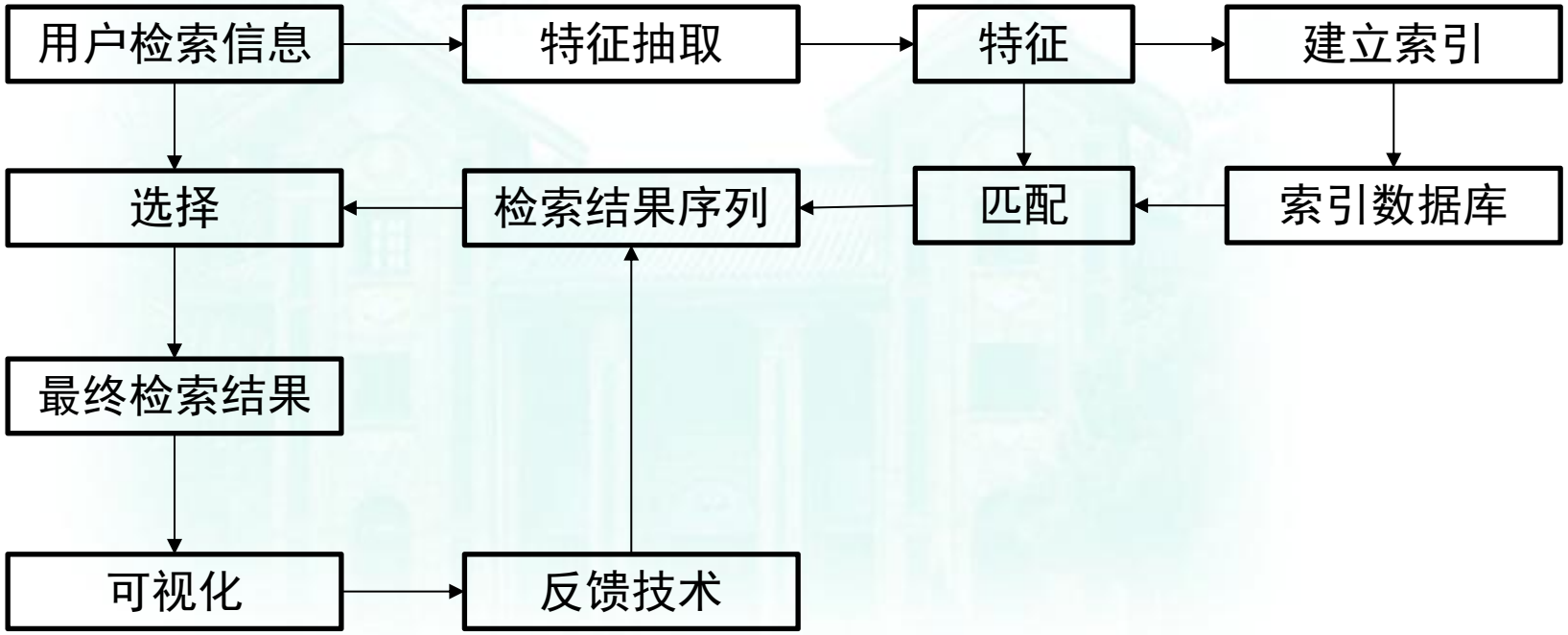
数字媒体检索技术

9.2 基于内容的媒体检索

9.2 基于内容的媒体检索

- 基于标签的媒体检索无法传达图像，视频或数据本身的数字化等媒体项目的真实性，异构数据组成的复合数据与用户识别所获得的语义内容相关联。
- 考虑到多媒体数据的这种固有特征，实现了基于内容的媒体检索。
- 与基础数据模型，感兴趣区域的先验知识以及表示查询的方案紧密相关。

检索技术的整体框图



检索的索引建立与匹配

- 特征抽取模型能抽取出检索对象具有代表性的特征，并用特征向量表示检索对象的特征，如检索对象A的特征向量为：

$$\vec{F}(A) = [x_1, x_2, \cdots, x_n]$$

- 使用特殊的数据结构（链表、树）将特征向量建立索引，方便查询。

- 匹配可以使用欧式距离、马氏距离等度量方法去比较检索对象与索引库中的特征的相似度。

匹配距离

- 假设多媒体对象采用*N*个特征来表示，两个多媒体对象分别表示为：

  向量*X*=(x_1, x_2, . . , x_N)，向量*Y*=(y_1, y_2, …, y_N)

- 欧氏距离

$$D_{euc1} = \sum_{i=1}^{N} | x_i - y_i | \qquad D_{euc2} = \sum_{i=1}^{N} (x_i - y_i)^2$$

- 马氏距离：C是特征向量的协方差矩阵

$$D_{mahal} = (\mathbf{X} - \mathbf{Y})^T C^{-1} (\mathbf{X} - \mathbf{Y})$$

- 其他方法

检索过程与分类过程的区别

- 以图像为例，分类是给定一幅测试图像，利用训练好的分类器判定它所属的类别，而分类器是利用带类别标签的训练数据训练出来的。

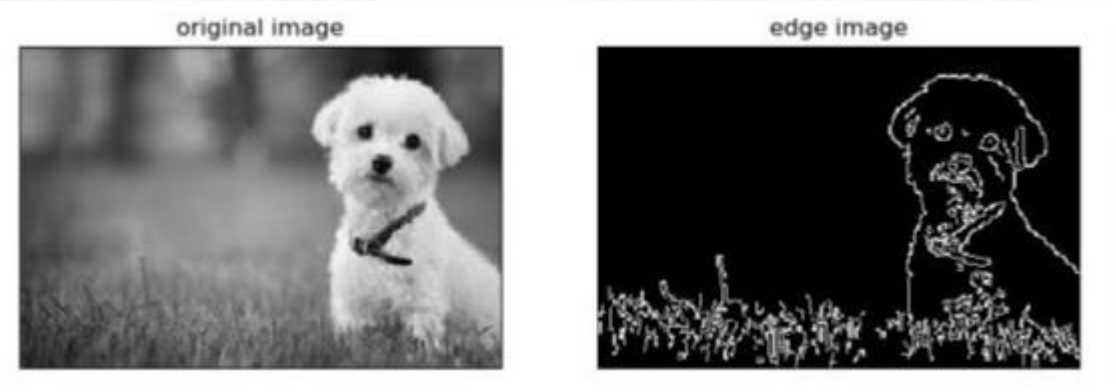
- 检索（如以图搜图）则是给定一幅查询图像，搜索与之相似（视觉或语义上）的图像。一般是提取图像特征后直接基于某种相似性（距离）度量标准计算查询图特征和数据库中图像特征之间的相似性，然后根据相似性大小排序输出结果。

9.2 基于内容的图片检索

- 图片：
  - 基于形状的检索
  - 基于色彩的检索
  - 基于空间关系的检索
  - 基于纹理的检索

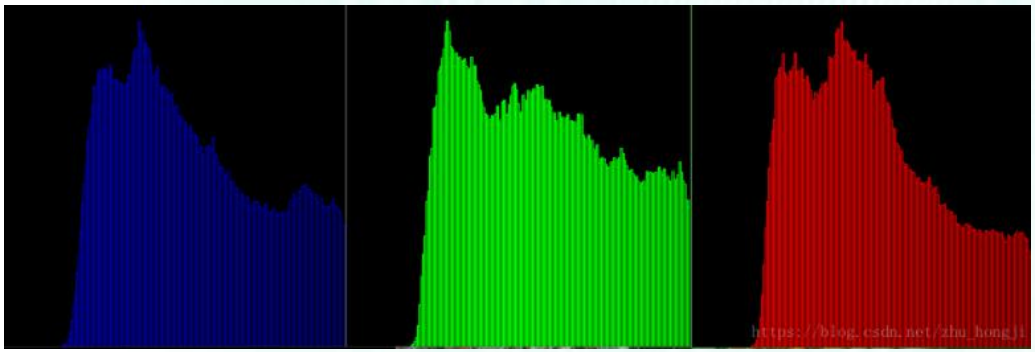
基于形状的检索

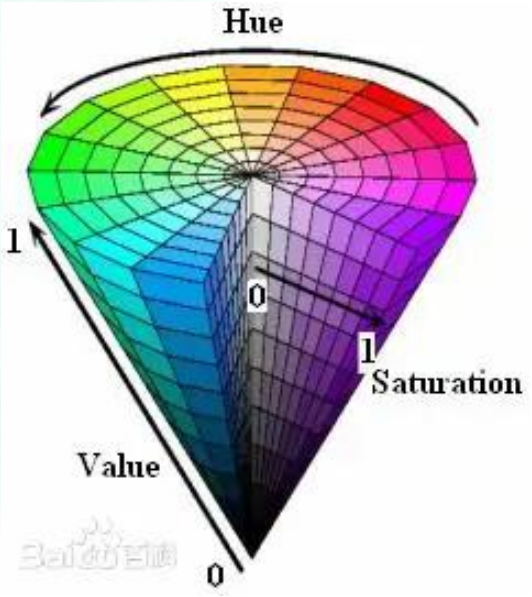
- 通过霍夫变换等边缘检测算法提取图像边缘轮廓，使用图像边缘轮廓与数据库中的图像资源相匹配。



基于颜色的检索

- 提取图像的RGB颜色累加直方图或者转换为其他颜色空间（HSV等）的直方图。



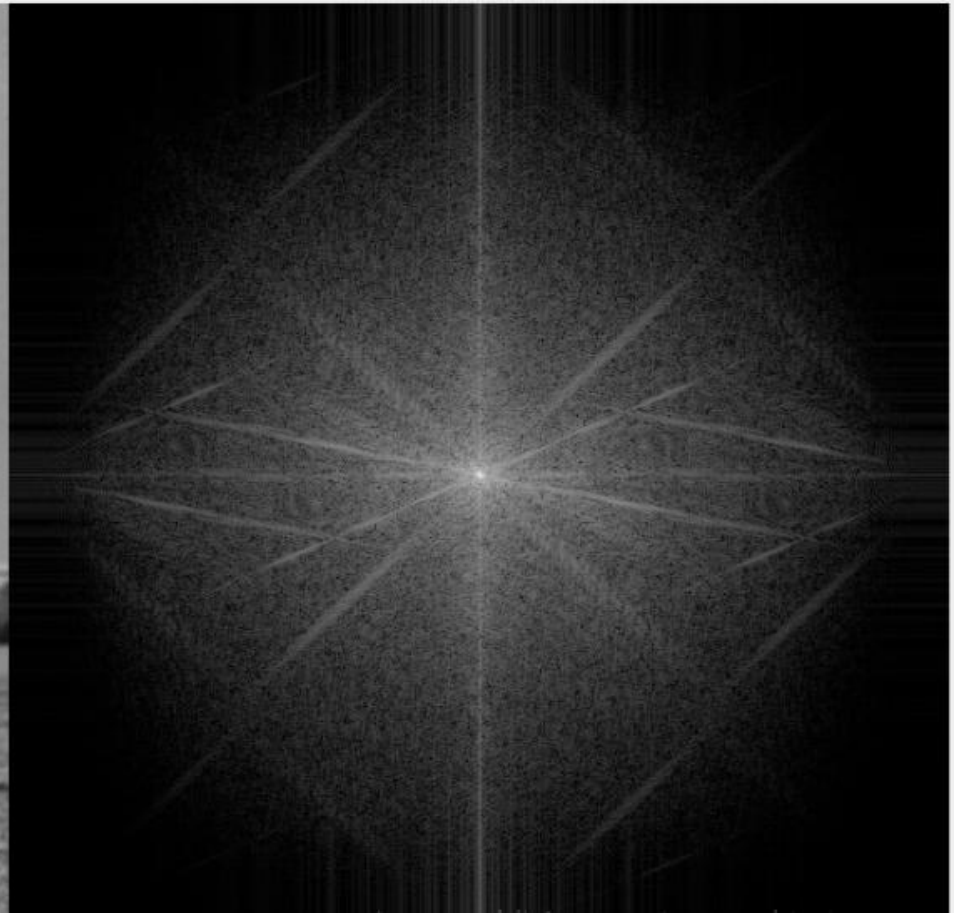


基于空间关系的检索

- 对图像进行分割，划分出图像中所包含的对象或颜色区域，然后根据这些区域提取图像特征，并建立索引。

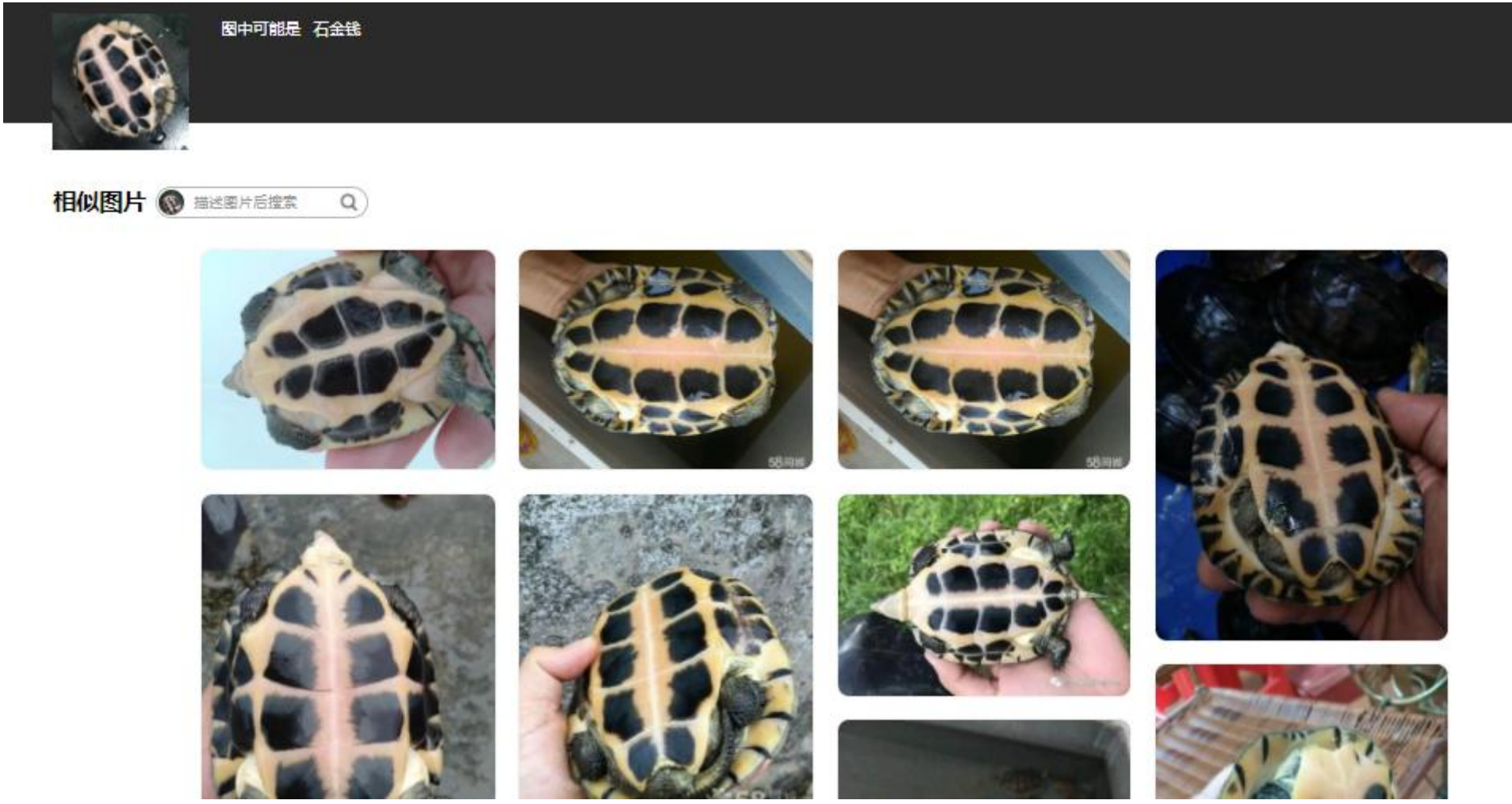
基于纹理的检索

- 使用共生矩阵、频谱分析等方法提取图像纹理进行检索。



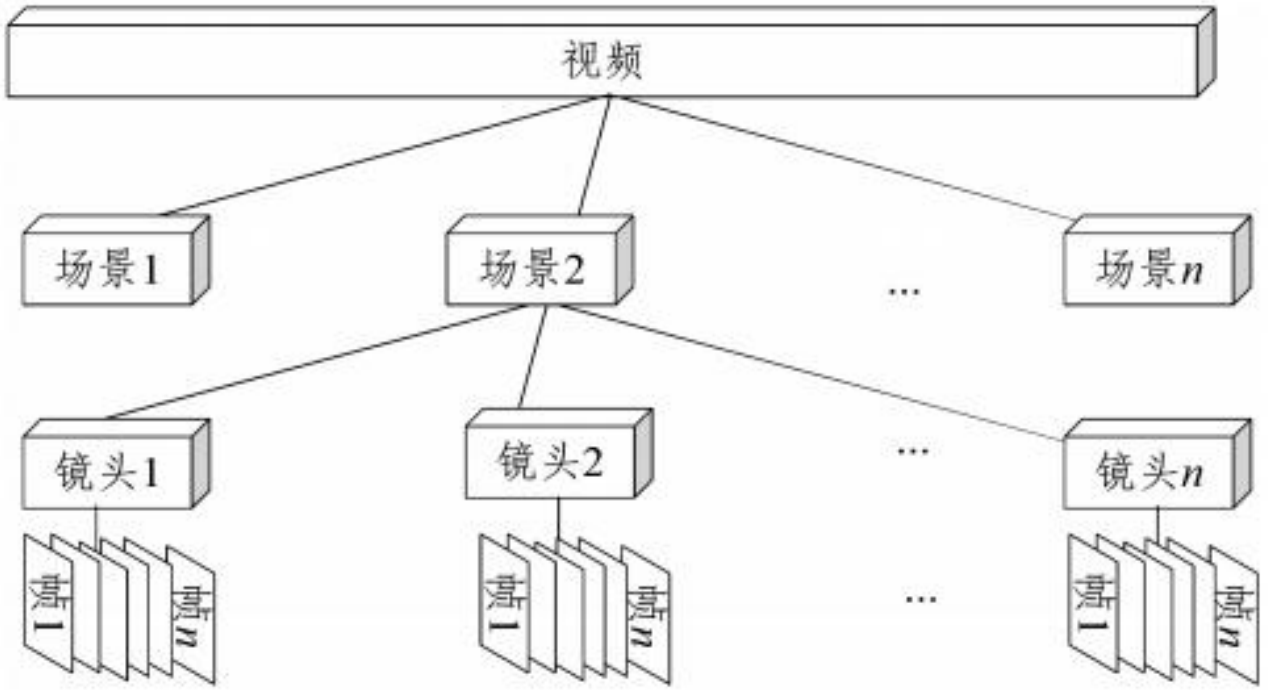
基于纹理的检索

- 以图搜图



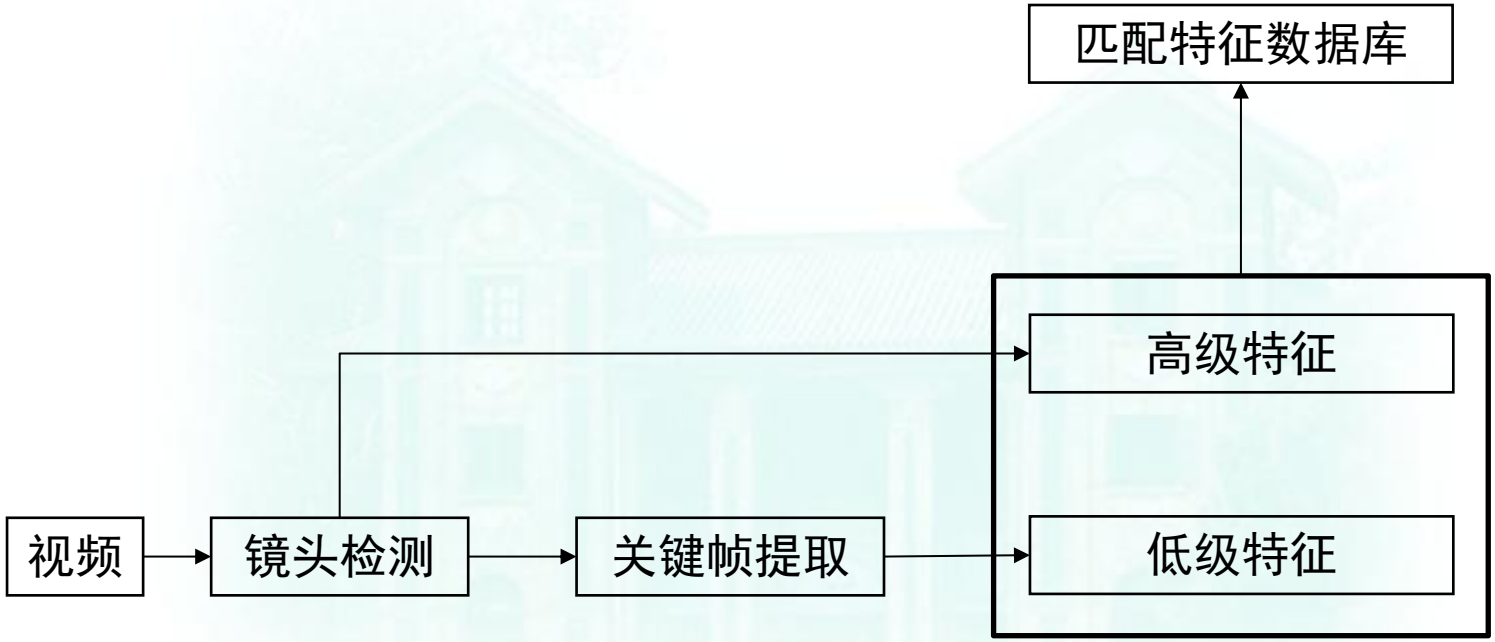
基于内容的视频检索

- 视频的层次结构



基于内容的视频检索

- 视频检索的流程



镜头检测

- 镜头检测用于将整个视频分割成多个镜头，镜头边界位置处的帧与属于下一镜头的连续帧之间存在视觉差异。
- 通常采用镜头边界检测的方法有：像素比较法、颜色直方图作差法等。

关键帧提取

- 由于同一镜头的帧存在冗余，因此选择一个或者多个最能反映镜头内容的帧作为关键帧来表示镜头，提取关键帧的关键在于选择最能反映镜头内容同时尽可能避免冗余的帧.

- 关键帧的提取方式可以分为人工指定、 参考帧、 聚类等.

特征提取

- 通常，视频数据特征的性质可以分为低级特征和高级特征。
- 低级特征可以来源于视频中提取的关键帧，或者视频中提取的物体。
- 高级特征也称为时间特征或者运动特征，是区别动态视频和静止图像的基本特征，它比静态关键帧特征更接近视频语义概念.

特征提取

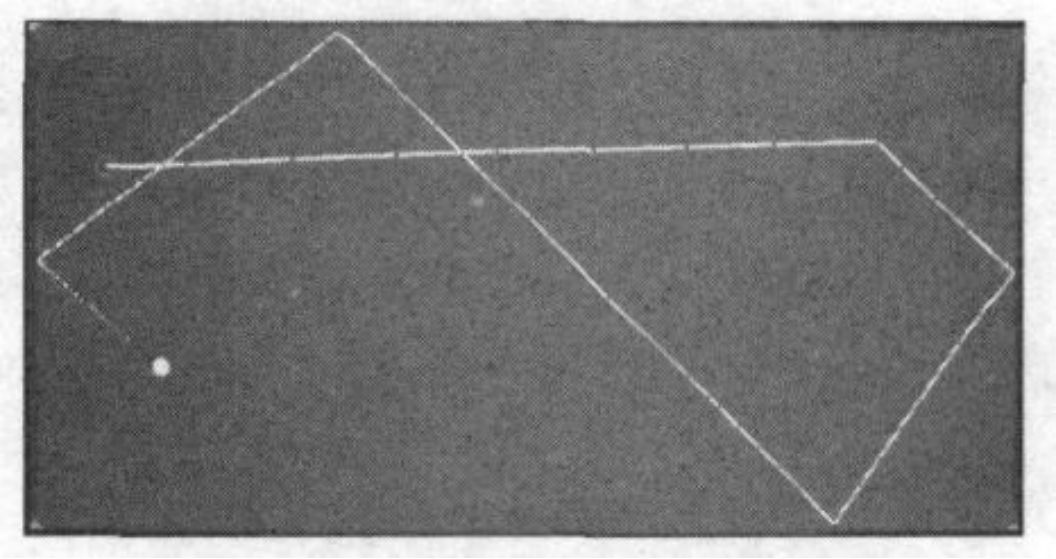
- 低级特征根据关键帧做图片特征的提取，如形状、颜色、纹理等。
- 高级特征根据视频中物体的运动特征进行提取，分为基于运动分割和基于运动轨迹两种方法。

运动分割

- 运动分割的目的是从序列图像中将变化区域从背景图像中提取出来。

运动轨迹

- 描述序列图像中物体的运动轨迹，并将其运动矢量编码为特征。



基于内容的音频检索

- 音频特征可以分为:听觉感知特征和听觉非感知特征(物理特性) 。
- 听觉感知特征包括音量 (loudness) 、音调 (pitch) 、音强 (brightness) 等。
- 听觉非感知特征包括对数倒频谱系数、线性预测系数等。

基于内容的音频检索

- 听歌识曲