计算机信息检索

第6章 多媒体检索(Multimedia IR)

几个问题

- 多媒体指的是什么
- 多媒体检索与文本检索的区别与联系
- 如何进行多媒体检索

提纲

1. 多媒体检索概述
2. 声音检索
3. 图像检索
4. 视频检索

多媒体定义

- 从定义上来说，多媒体也包括文本这种媒体形式。
- 但是，通常上的多媒体往往特指除去“文本”以后的各种媒体。
- 本章将的多媒体检索中的多媒体就指的是后面这个概念。

多媒体定义

- 多媒体对象
- 网上存在大量多媒体文档
  - 声音： mp3/wav/rm…
  - 图片： jpg/bmp/gif/tiff/…
  - 动画： swf/gif…
  - 图形： (矢量图形文件)dwg/dxf/3ds…
  - 视频： mov/wmv/mpeg/mpg/rm…

A picture is worth a thousand words !

多媒体定义

- **多媒体文档非常普遍**
  - 计算机硬件不断升级
  - 网络带宽不断扩大
  - 摄录设备日益普及
    - DC/DV/Web cam
  - 多媒体制作日益平民化
  - 传播渠道日益广泛

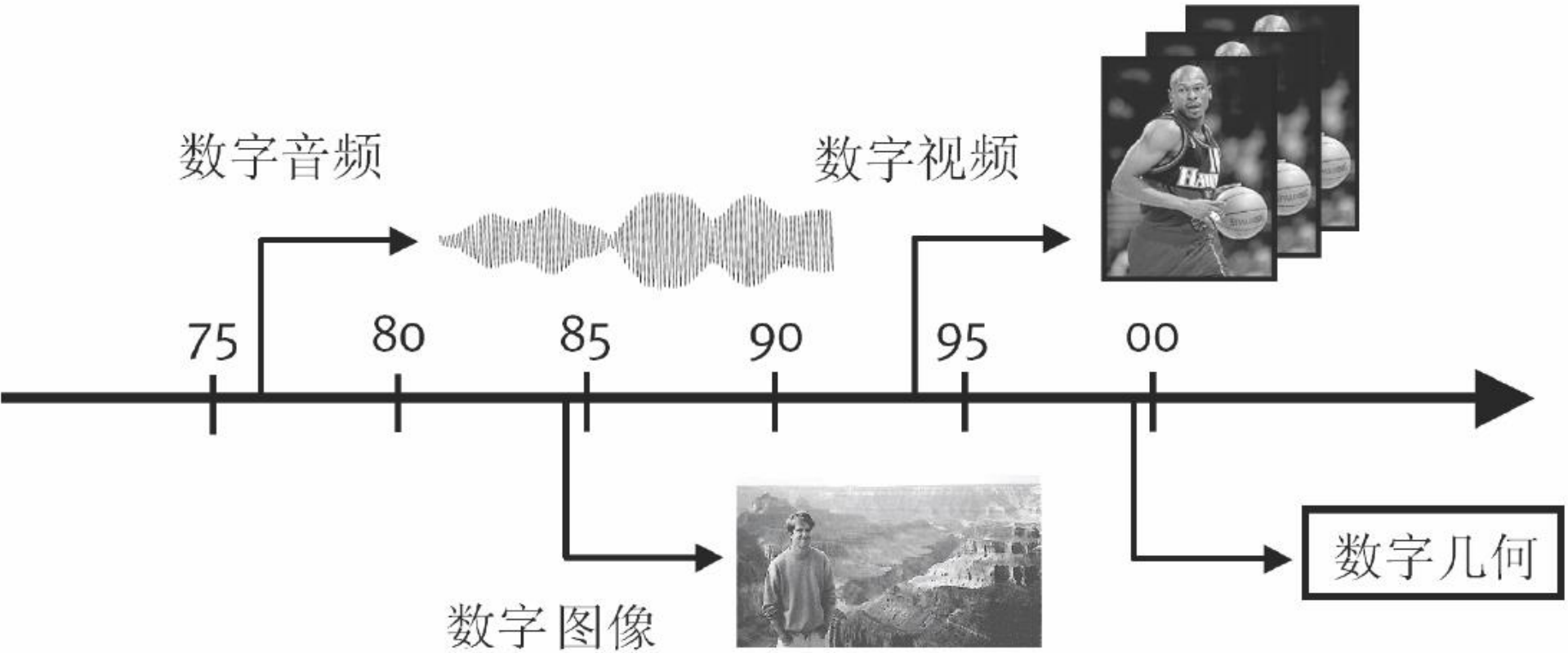
多媒体文档更具娱乐性

- 馒头血案　　小胖　　**Youtube**

多媒体检索非常困难

- 对同一主题，多媒体表达千差万别
- 多媒体对象具有十分复杂的特征，进行特征表示比较困难，对多媒体对象的理解就更困难
- 用户的检索需求也非常复杂，有时是基于低级特征、有些是基于元数据文字描述、有些是基于高级语义特征。

多媒体检索发展历史



多媒体检索成为竞争焦点

- 以搜索引擎为代表的文本检索已经深入人心，得到了用户的认可。
- 而多媒体检索却由于技术上的难度目前在应用上并没取得突破，离用户的要求还有较大的距离。
- 各大公司投入很大力量进行多媒体检索的研发。

多媒体检索的方法(1)

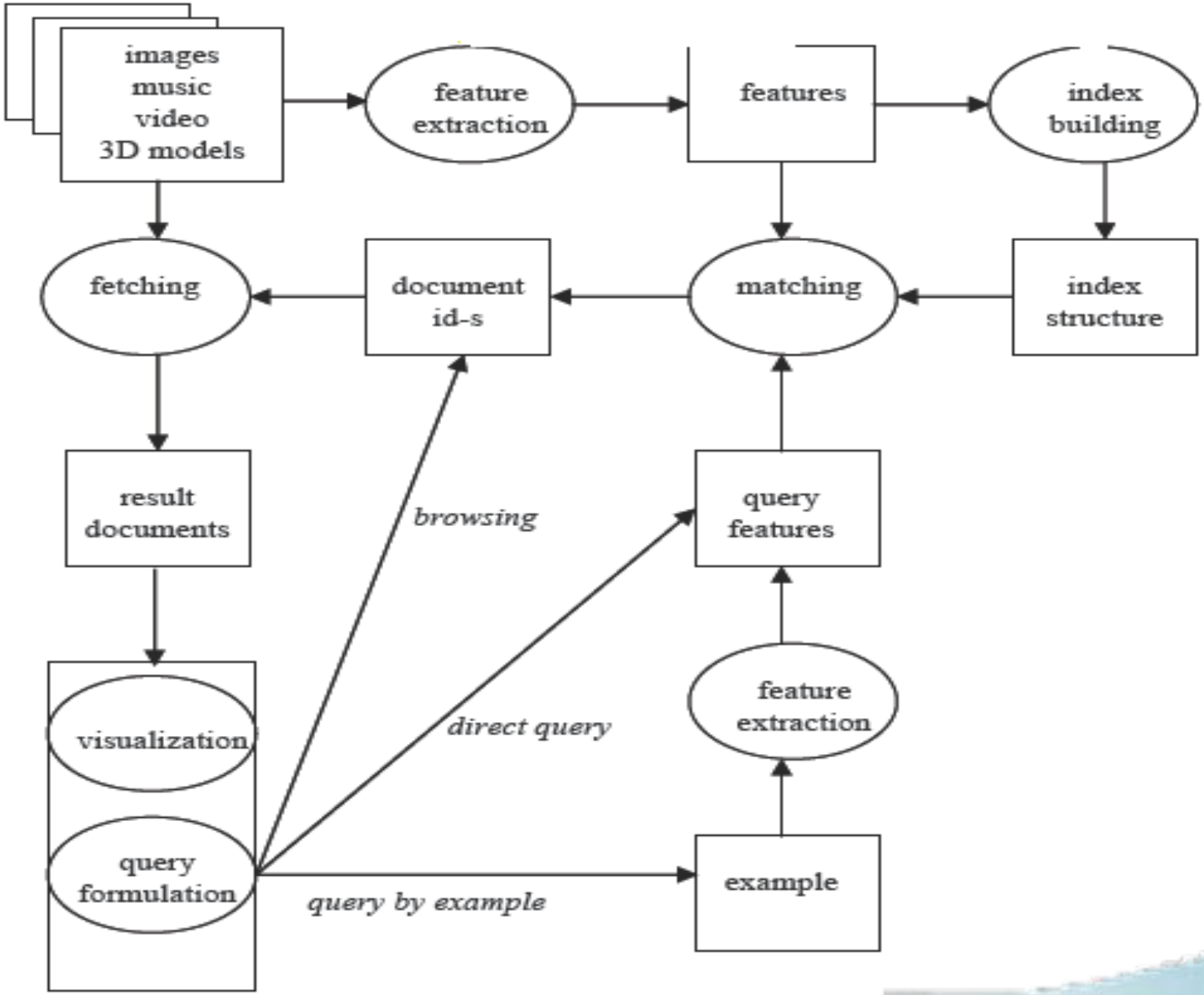
基于关键词检索的方法

- 人工标注：对多媒体对象进行手工标注，可标注元数据(作者、标题、日期等)或者内容数据(内容关键词)。如WEB2.0中提交多媒体对象时的标签(tag)数据就是标注文本。
- 自动抽取：
  - 在多媒体对象周围抽取能够表示对象的文本数据用于标注。如在WEB中通过图片周围的文字来描述图片。
  - 在视频中抽取字幕、对话，从音频中抽取语音，从图片中识别文字等等。

多媒体检索的方法(2)

- **基于内容的方法(Content Based Retrieval，CBR)**
  - 从多媒体对象的内容出发，抽取它们的特征并进行特征表示，在特征层面上进行相似度计算，得到检索结果。
    - 如：基于颜色或形状的图像检索、哼一句歌找整支歌曲、基于概念的检索(如：检索有关“日出”的图片)
  - CBR是当前大多数研究所关注的方法。

多媒体检索的一般框架


images
music
video
3D models
feature extraction
features
index building
fetching
document id-s
matching
index structure
result documents
browsing
query features
visualization
direct query
feature extraction
query formulation
query by example
example

多媒体对象中的特征

- **视觉类媒体的特征：**
  - 颜色、形状、纹理、空间约束、运动、对象(如太阳)、场景、语义(如日出)等等
- **听觉类媒体的特征：**
  - 音调、音量、音色、旋律、和谐度、语义(如爆炸声)等

相似度计算

- 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=(x1,x2,..,xN)，向量Y=(y1,y2,…,yN)
- 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

QuerybyExample(基于样例的查询)

query:

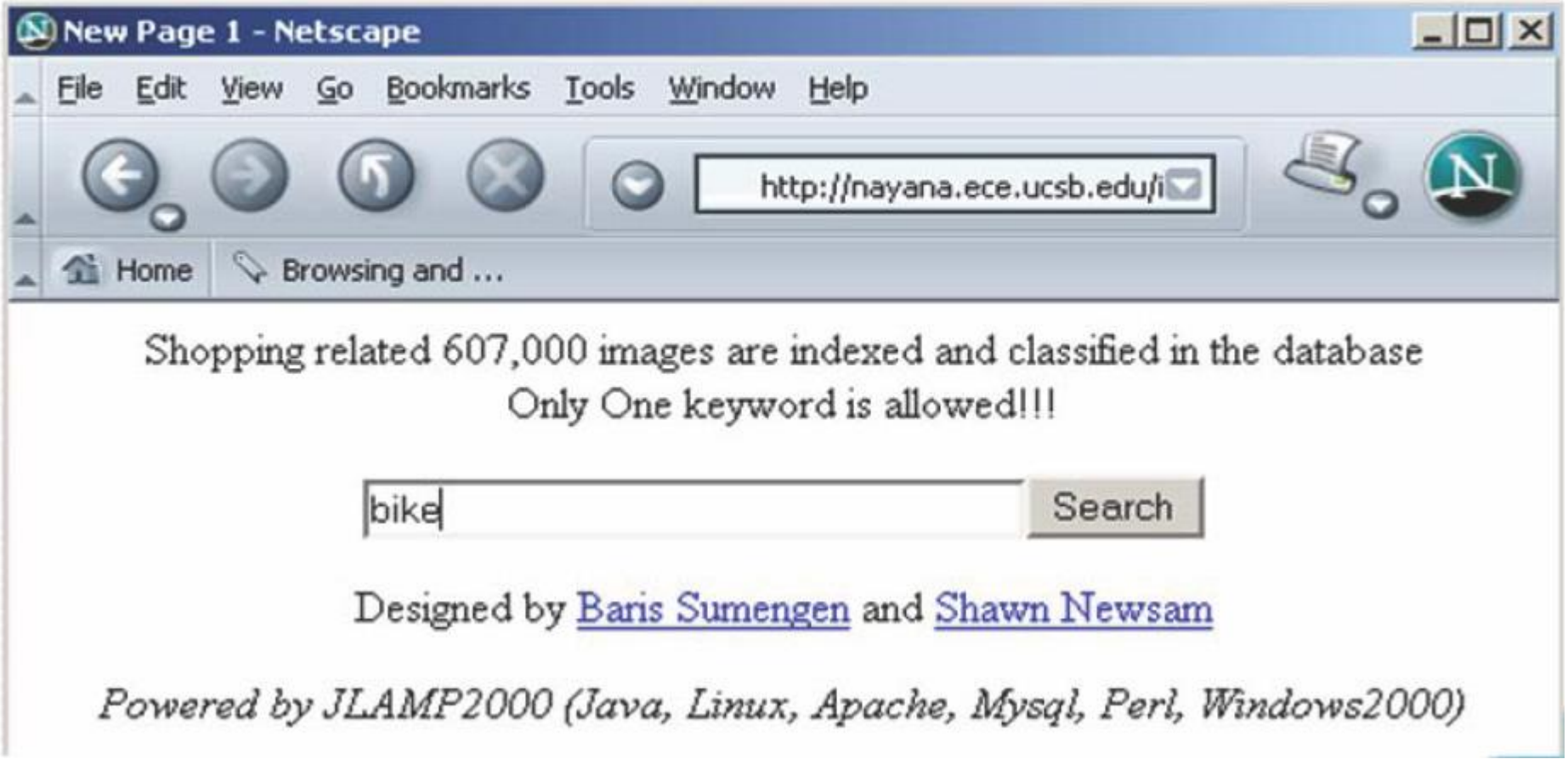
target:

QuerybySketch(基于草图的查询)

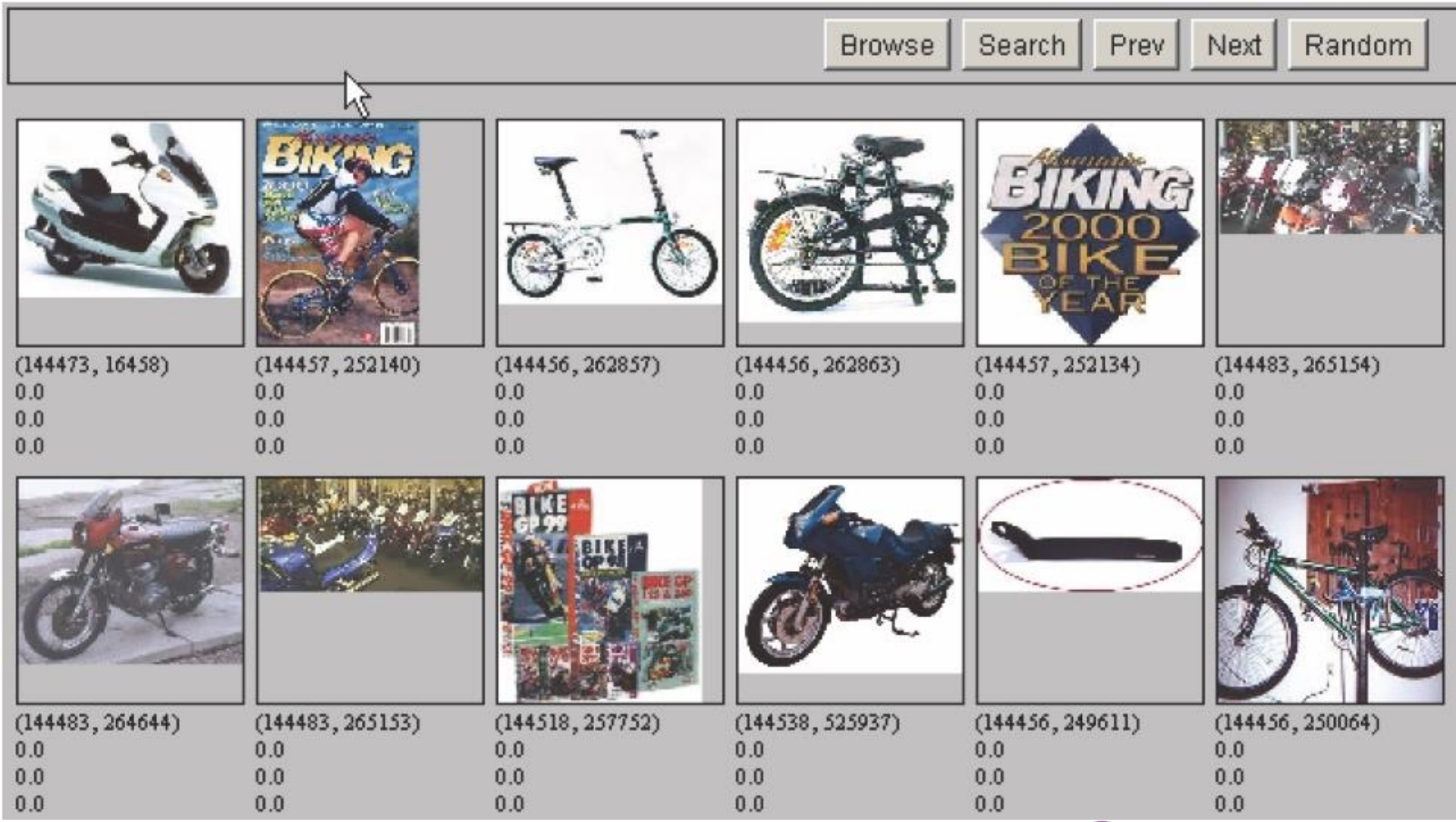
多媒体检索中的相关反馈

Image Search Engine

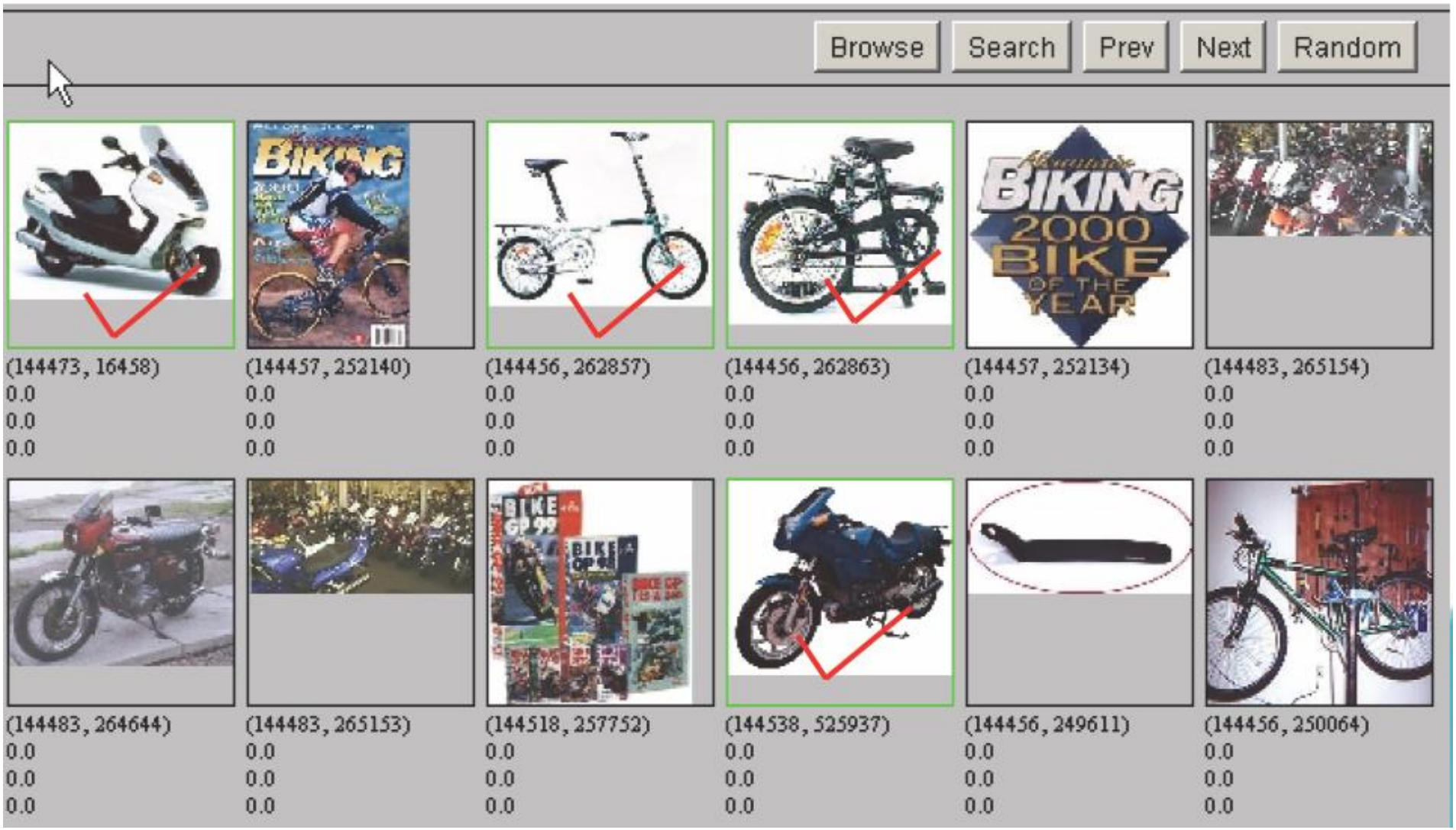
http://nayana.ece.ucsb.edu/imsearch/imsearch.html



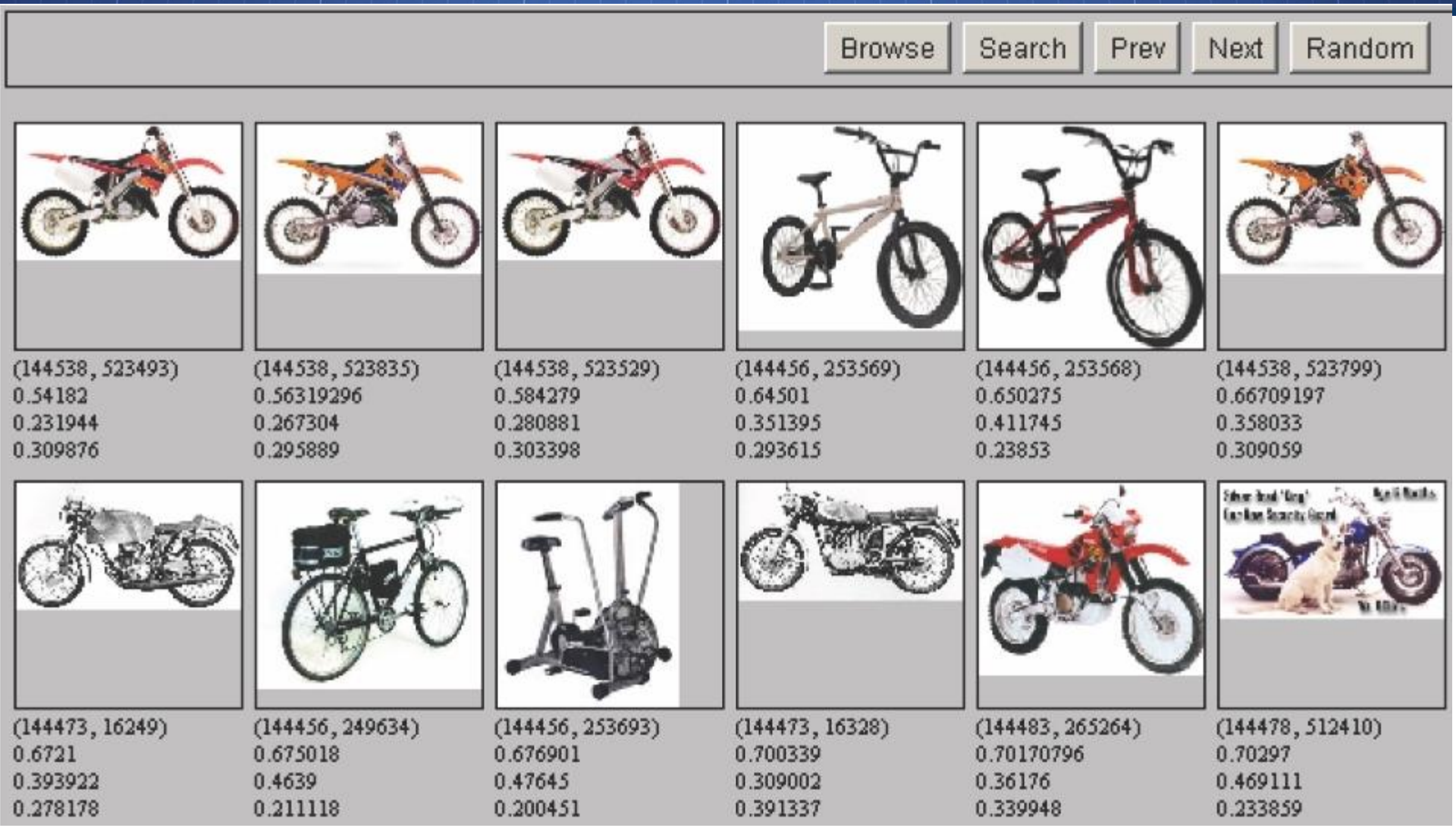
初始结果



(用户)相关反馈



再次检索的结果



一些多媒体检索的应用

- Logo retrieval
- CAD searching
- Product catalogues
- Museum collections
- Photo archives
- Music selection
- Medical imaging
- Crime investigation, law enforcement
- Video searching
- Encyclopedia search
- Copyright protection

跨媒体检索(Cross-media retrieval)

- **是指查询和检索对象分属于不同媒体表达形式的检索，如：利用天鹅的叫声去检索天鹅的图片。**
- **跨媒体检索通常还会涉及两个意思：**
  - 检索结果的呈现上，可以采用多种媒体形式共同表达
  - 利用多模态(multimodal)信息弥补单模态信息的不足：如视频中通常也包含文字和音频流，可以利用它们的综合信息为检索服务。

提纲

1. 多媒体检索概述
2. 声音检索
3. 图像检索
4. 视频检索

音频(audio)

- 音频(声音)经过模拟设备记录或再生，成为模拟音频，再经数字化成为数字音频。
- 数字音频的主要规格为：采样率(sampling rate)及每个样本的位数(bits per sample)。
- 我们能够听见的音频频率范围是60Hz～20kHz，其中语音(speech)大约分布在300Hz～4kHz之内，而音乐(music)和其他自然声响是全范围分布的。

音频规格

- 采样率：对模拟声音采样时，每秒钟取的样本数目。数字化时的采样率必须高于信号带宽的2倍，才能正确恢复信号。
- 每个样本的位数：对每个样本的表示所采用的位数，如8或16。位数越大，声音的表示越精确，所需要的存储空间也越大。
- 以普通CD为例，通常是采用44.1kHZ(1k=1024)的采样率，每个样本采用16位表示，则1秒钟需要705.6kb表示。

音频中的特征层次

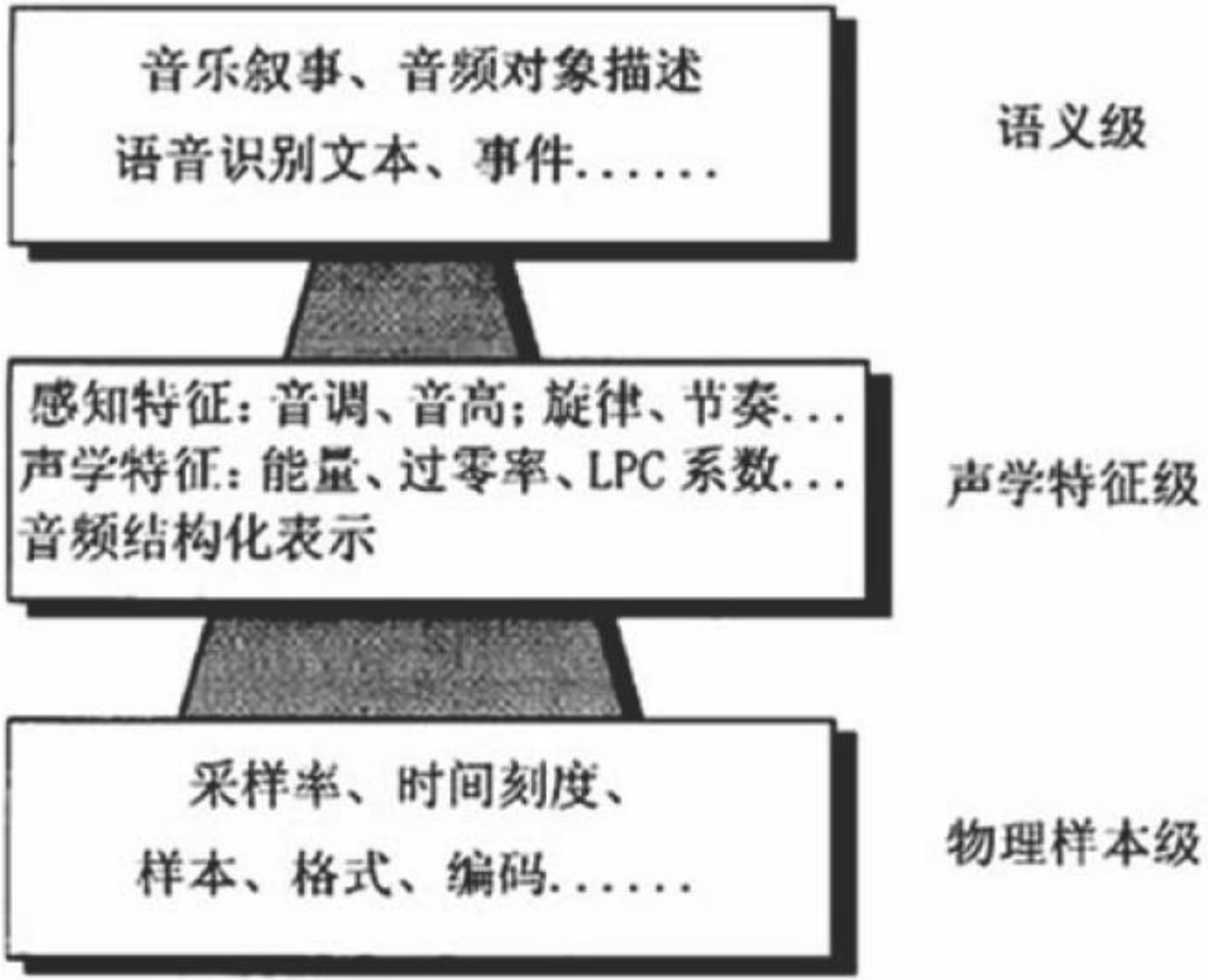


图　音频内容分层描述模型

查询形式(1)

- 样例：用户选择一个声音例子表达其查询要求，查找出与该声音在某些特征方面相似的所有声音。如查询与飞机的轰鸣声相似的所有声音
- 直喻：通过选择一些声学/感知物理特性来描述查询要求，如亮度、音调和音量等。

查询形式(2)

- 拟声：发出与要查找的声音性质相似的声音来表达查询要求。如用户可以发出嗡嗡声来查找蜜蜂或电气嘈杂声。
- 主观特征：用个人的描述语言来描述声音。这需要训练系统理解这些描述术语的含义，如用户可能要寻找“欢快”的声音。
- 浏览：基于分类目录或音频的结构进行浏览

语音检索(Speech Retrieval)

- **主要利用语音识别(Speech Recognition)技术，从语音中获取全部文本或者关键文本、或者辨别说话人。**
  - 抽取全部文本，根据文本建立索引，进行文本检索。
  - 抽取关键词，比如抽取“进球”来标识进球语音。
  - 辨别说话人，比如通过辨别说话人的变化对语音进行分割。

普通音频检索

- 以波形声音为对象的检索，这里的音频可以是汽车发动机声、雨声、鸟叫声，也可以是语音和音乐等，这些音频都统一用声学特征来检索
  - 音频分割
  - 音频训练及分类
  - 基于听觉特征检索

音乐检索

- 以音乐为中心的检索，利用音乐的音符和旋律等音乐特性来检索。如检索乐器、声乐作品等。
  - 基于样例检索
  - 基于哼唱曲调来检索

音乐的语义检索

音乐的语义检索

有一首英文歌吹口哨?

1 人赞同了该回答

是不是 Home - Edward Sharpe & The Magnetic Zeros，虽然只是前段和中间有口哨，但是类型挺像你说的。
你还记得那首歌是男声还是女声还是合唱，节奏舒缓程度什么的。。
楼主你是不是消失了。。。

我干脆给你链接自己去听。。 口哨前奏，一次听个够_Kiri精选集
以及 口哨悠扬【蓝色出品 史上最强】
以及 歌曲: 最好听的音乐试听，mp3下载口哨

编辑于 2015-07-12

▲ 1 ▼ 1 条评论 分享 收藏 感谢

生帆子wide
和气生财、

12 人赞同了该回答

whistle

发布于 2014-10-11

▲ 12 ▼ 3 条评论 分享 收藏 感谢

小爷
微信平台：#奇怪又浪漫的古典诗人# StrangeWorldForU

南京理工大学
NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

提纲

1. 多媒体检索概述
2. 声音检索
3. 图像检索
4. 视频检索

图像(image)

- 二维材料经扫描器扫描、拍照或编辑产生数字化图像。图像的主要规格包括分辨率、颜色表示位数、存储格式、压缩手段等等。
- 图像包括：照片(photo)、图片(picture)、位图(bitmap)、电脑绘图(graphics)、视频中的帧(frame)。

图像规格

- 分辨率(resolution)：图像在横方向和纵方向的像素个数，用“宽*高”表示。如1024*768。
- 每个像素的表示位数：每个像素是单色或者彩色。8位表示：0~255表示单色的灰度值。24位表示：每8位分别表示红绿蓝3原色。
- 不压缩情况下，一幅1024*768的24位彩色表示图像占用的存储空间为768*3=2304KB
- 存储格式、压缩方法：gif/jpg (Joint Photographic Experts Group)/ bmp/tiff等等

图像视觉特征

- 颜色(color)：图像的颜色分布。
- 纹理(texture)：纹理是指图像局部不规则的而宏观上有规律的特征，人们区分纹理主要使用粗糙性和方向性两个方面。
- 形状(shape)：物体的边界特征或者主要轮廓

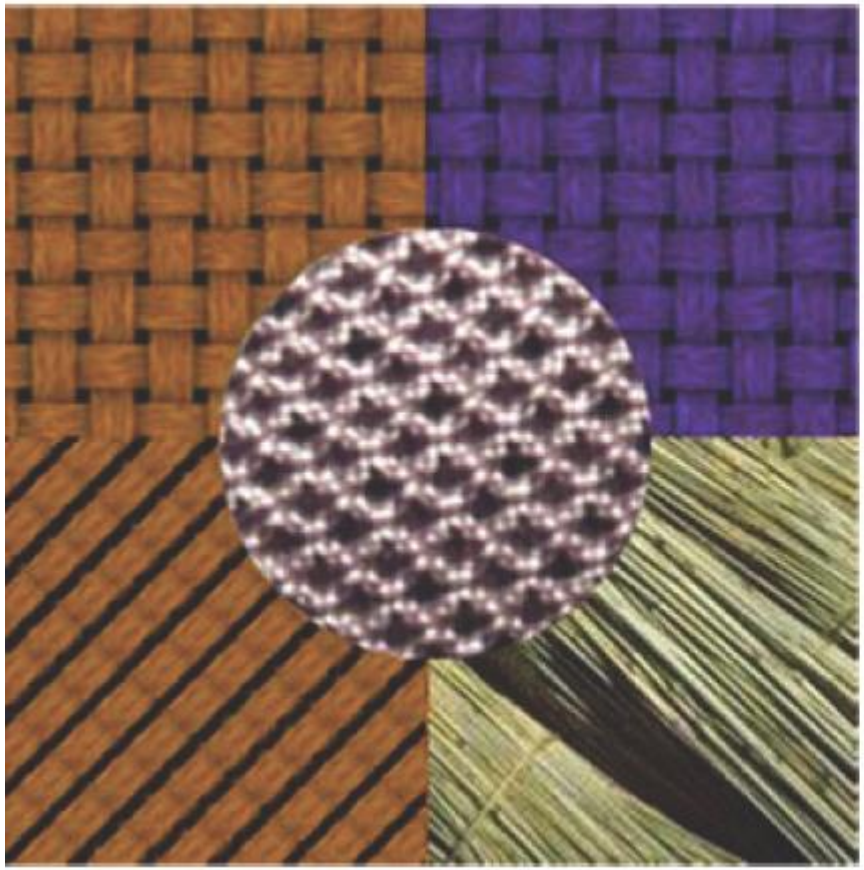
颜色特征

- 统计主要颜色的分布

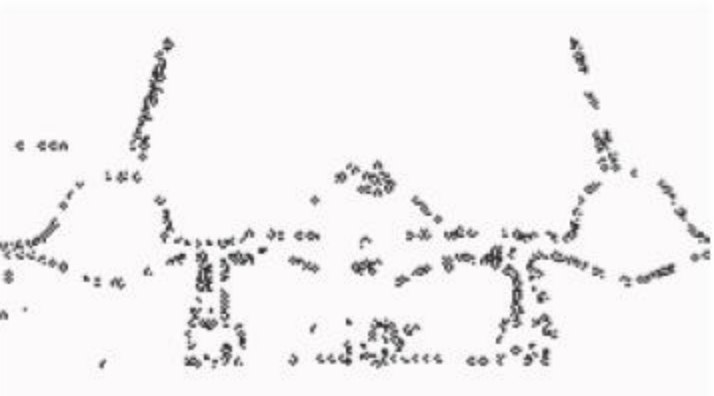
纹理特征

- 某颜色或密度模式的改变

纹理的分割



形状特征



查询形式

- 样例：根据库中或者库外已有图像或者人工绘制的图像进行检索。比如通过输入一个红色圆形物体来检索相似的图像。
- 绘图：手工绘制草图用于检索。如通过勾画衣服形状对服装设计图进行检索。
- 属性说明方式：指定特征进行检索。如通过限定人的脸形、五官特征从人脸库中进行检索。
- 浏览方式：按类别或者库结构进行浏览。

基于视觉特征的检索

- 基于颜色特征进行检索：检索出与用户颜色要求相似的图像。在检索中，颜色空间常常不采用RGB方法，而是采用HSV方法(hue-色调，saturation-饱和度，value-亮度)
- 基于纹理特征的检索：检索出与用户纹理要求相似的图像。
- 基于形状特征的检索：检索出与用户形状要求相似的图像。主要通过主要边界特征或轮廓特征来实现。

基于对象和区域特征的检索

- 基于全局特征：全局特征包括图像总的色调、颜色统计分布、图像的一般属性(如图像中的对象数目、总面积等等)和视觉特征。
- 基于局部特征：局部对象的颜色、纹理或形状，对象在空间的约束逻辑关系(方向、邻接或包含)。

基于综合特征的检索

- 将不同侧面的特征综合起来进行图像的检索。如将图像的客观属性(如：作者、时间)、主观属性(如：人的胖瘦)或者语义属性(如：日出)结合在一块进行检索。

文字型图像的检索

- 文字型图像(textual image)：通过对书面文本进行扫描得到的图像。
- 通过OCR系统识别图像中的文本，基于文本进行检索。

提纲

1. 多媒体检索概述
2. 声音检索
3. 图像检索
4. 视频检索

视频(Video)

- 主要通过视频采集卡从播放画面中采集加工而成。可以看成是在普通图像上增加了时间维度。主要的规格包括：分辨率、每秒播放帧数、压缩方法等。
- 常见的视频格式：.dat、.mov、.rm、wmv、mpg、mpeg等等
- 每秒播放帧数：电视是30帧，电影为24帧，对人的感觉而言，至少要每秒12帧以上。
- 压缩方法：MPEG (Motion Picture Experts Group)、国内AVS

视频中的特征层次(1)

视频级	概要	
	语义	
	一般属性	
场景级	场景关键帧	(静态特征)
	场景关键对象及其特征	(动态特征)
	其他运动特征	(动态特征)
镜头级	镜头关键帧	(静态特征)
	镜头关键对象及其特征	(动态特征)
	其他运动特征	(动态特征)
帧级	静态图像特征	(静态特征)

视频中的特征层次(2)

- 帧(Frame)：每个帧可以看成一幅静态图像。
- 镜头(Shot)：由连续的帧组成的一个基本拍摄操作单元。镜头可以通过关键帧表示，摄像机操作引起的镜头运动特征也是视频检索中重要的特征内容。
- 场景(Scene)：由连续的多个内容相似的镜头组成的一个有意义的单元。场景关键帧可以由镜头关键帧组合而成。关键对象也可以组合。
- 视频级特征：完整的视频故事或者节目，包含视频的概要、语义和一般属性的描述。

视频的分析及检索

- 镜头边界检测(镜头分割)
- 关键帧提取
- 镜头聚类及场景识别
- 视频摘要
- 视频的浏览
- 视频的检索

视频的浏览

- 基于基本结构的浏览：按照视频层次结构找到视频单元进行播放或者浏览
- 基于事件和故事进行浏览：按照事件或者故事的发生进行浏览。

视频的检索

- 基于关键帧的检索：类似于图像检索的方法，利用全部和局部的图像特征进行检索。
- 基于运动特征的检索：基于摄像机运动或者像素运动特征的检索。
- 基于视频对象的检索：利用视频对象的特性，从库中检索出包含相关视频对象的所有场景或者镜头。

小结

1. 多媒体检索概述
2. 声音检索
3. 图像检索
4. 视频检索

The End