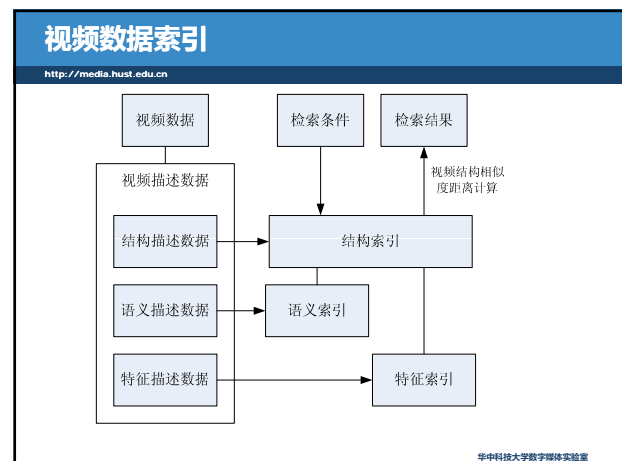
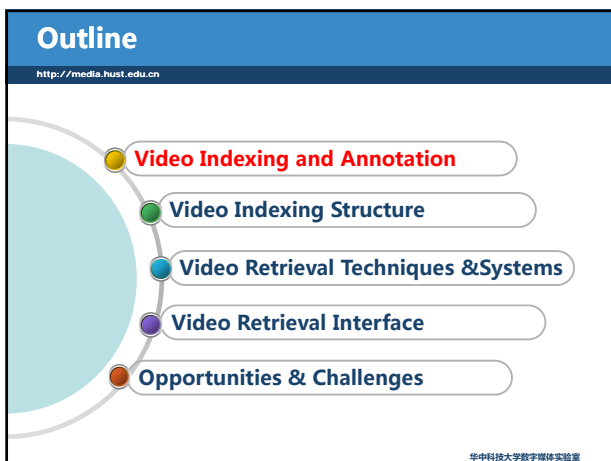
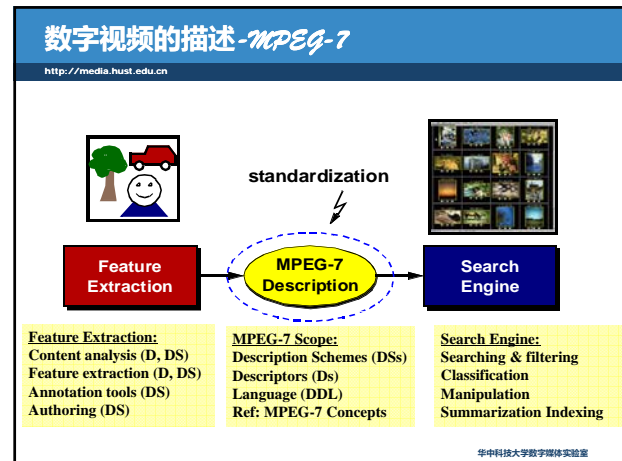


基于内容的 视频索引与检索

主讲：于俊清
<http://media.hust.edu.cn>

华中科技大学数字媒体实验室



视频索引的分类

<http://media.hust.edu.cn>

- ❖ 从检索方法上可以分为五种类型的索引：
 - 目录索引（标题、类别等）
 - 结构索引（镜头、场景等）
 - 内容索引（场景中的角色、运动目标等）

华中科技大学数字媒体实验室

结构索引

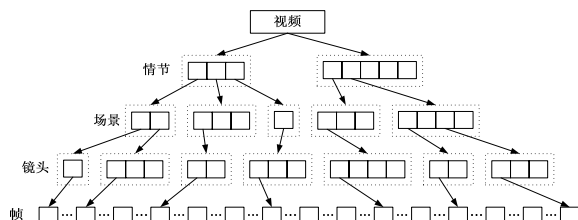
<http://media.hust.edu.cn>

- ❖ 结构索引是对视频内容的结构单元信息，例如关键帧、镜头信息等进行索引
- ❖ 结构索引为视频内容单元的浏览提供方便，它主要通过视频数据内容单元自身的层次结构来实现
- ❖ 结构索引在检索中主要用于返回用户所需要的视频内容单元集

华中科技大学数字媒体实验室

结构索引（续）

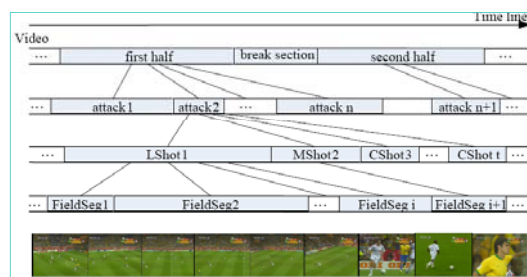
<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

结构索引（续）

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

视频索引的分类

<http://media.hust.edu.cn>

❖ 从索引的生成方式出发，可以将索引分为：

- 手工索引
- 半自动索引
- 自动索引



华中科技大学数字媒体实验室

视频索引的分类

<http://media.hust.edu.cn>

❖ 从选取的索引内容出发，还可以把视频索引分成三类：

- 基于注释的索引(Annotation-based Indexing)
- 基于特征的索引(Feature-based Indexing)
- 基于特定领域的索引(Domain-specific Indexing)

华中科技大学数字媒体实验室

基于注释的索引

<http://media.hust.edu.cn>

- ❖ 所谓“**视频注释**”（annotation），是指用一些描述性的信息（如文字、声音或图形）来表述所指向的视频片段
- ❖ 视频注释提供了根据语义内容而不仅仅是颜色分布之类的声像内容来访问视频数据
- ❖ 完全自动的视频注释在可以预见的时期内是无法实现的，因此目前的视频注释一般是采用人机交互方式进行

华中科技大学数字媒体实验室

注释语言的选择

<http://media.hust.edu.cn>

- ❖ **文字注释**：分为语句注释和关键词注释
- ❖ **图标注释**：如电视台台标
- ❖ **源注释**：利用数字摄影机在视频数据流中加入相关信息作为视频注释的依据，如摄像机运动

华中科技大学数字媒体实验室

基于特征的索引

<http://media.hust.edu.cn>

- ❖ 视频数据包含**语法内容**和**语义内容**
- ❖ 目前机器视觉还不能自动识别语义内容，但可自动识别语法内容
- ❖ 基于特征的索引模式正是利用可自动识别的语法内容建立部分视频索引，最大限度地减少手工操作的工作量
- ❖ 基于特征索引技术的目标是全自动索引

华中科技大学数字媒体实验室

基于特征的索引（续）

<http://media.hust.edu.cn>

- ❖ 可用于索引的视觉特征包括：颜色、纹理、形状、空间关系、摄像机运动等
- ❖ 视频流中的文字信息
- ❖ 视频流中的音频特征，如静音、短时能量、和谐度、频谱分布等

华中科技大学数字媒体实验室

基于特征的索引（续）

<http://media.hust.edu.cn>

- ❖ 视频特征，尤其是低层特征的领域相关性小，大部分特征可以从视频中直接提取
- ❖ 特征在检索时可以利用不同的相似度距离函数度量
- ❖ 一种特征仅能描述视频的某一方面的内容，实现特征检索需要多个特征

华中科技大学数字媒体实验室

基于元特征数据的索引

<http://media.hust.edu.cn>

- ❖ **元（Meta）特征数据**是关于数据的数据
- ❖ 元数据特征并不是表示直接从视频数据获取的信息，而是描述视频数据自身的一些信息
- ❖ 典型的例子包括镜头拍摄的日期、节目制作的方式等

华中科技大学数字媒体实验室

基于元特征数据的索引（续）

<http://media.hust.edu.cn>

- ❖ 元特征数据常借助文本的形式嵌入到视频流中，如播音员姓名、演员表、新闻标题等
- ❖ 这些文本内容也提供了视频节目的元特征
- ❖ 如果可以将它们从视频中孤立（分割）出来，并利用光学特征识别技术识别出来，就可以得到许多有关视频的语法和语义内容

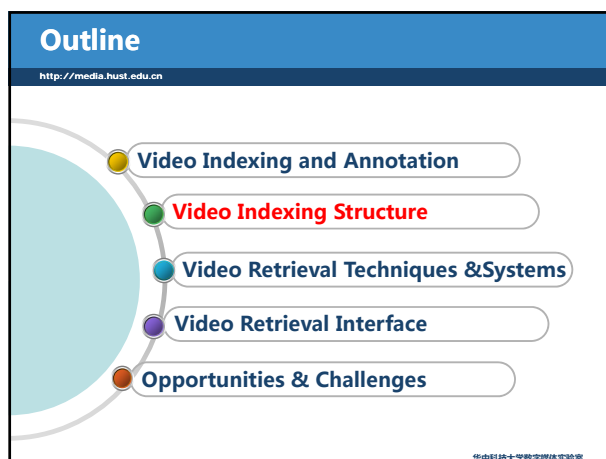
华中科技大学数字媒体实验室

基于领域知识的索引

<http://media.hust.edu.cn>

- ❖ 电视新闻视频索引
- ❖ 体育视频索引
- ❖ 故事片视频索引

华中科技大学数字媒体实验室



高维数据索引结构基础

<http://media.hust.edu.cn>

- ❖ 高维数据索引的目的
- ❖ 高维数据索引的方法
- ❖ 向量空间与度量空间
- ❖ 常见的高维索引结构

华中科技大学数字媒体实验室

高维数据索引的目的

<http://media.hust.edu.cn>

- ❖ 降低磁盘的I/O操作次数
- ❖ 减少距离计算的次数，降低检索算法的复杂程度

华中科技大学数字媒体实验室

高维数据索引的方法

<http://media.hust.edu.cn>

- ❖ 降维，即将高维数据索引的维度减少，用维度减少后的数据进行相似度检索得到粗略的结果集
- ❖ 然后对结果集中数据按照原来的维度进行相似度计算，最后得到精确的结果集

华中科技大学数字媒体实验室

高维数据索引的方法（续）

<http://media.hust.edu.cn>

- ❖ 利用高维数据在空间的分布，将数据库中的数据分散到若干个空间区域，建立高维索引结构
- ❖ 检索时通过检索数据与高维索引结构，对所有的空间区域进行判断，删除不可能包含结果集数据的空间区域
- ❖ 然后对剩下区域中的数据逐一计算相似度，得到最后的结果集

华中科技大学数字媒体实验室

向量空间与度量空间

<http://media.hust.edu.cn>

❖ 向量空间

- 距离值采用欧式距离度量
- 数据的分布有空间特性（坐标值）
- 检索算法的代价主要是磁盘I/O操作的次数
- 典型的索引结构：R系列树，B系列树

华中科技大学数字媒体实验室

向量空间与度量空间

<http://media.hust.edu.cn>

❖ 度量空间

- 距离值采用相似度距离函数度量
 - 非负性： $d(x,y) \geq 0$ 且 $d(x,x)=0$
 - 对称性： $d(x,y)=d(y,x)$
 - 三角不等性： $d(x,y) \leq d(x,z)+d(z,y)$
- 检索算法的代价主要是距离函数的计算次数
- 典型索引结构：M系列树，VA-FILE

华中科技大学数字媒体实验室

传统的索引方法

<http://media.hust.edu.cn>

❖ 哈希表

- 数值的精确匹配
- 不能进行范围查询

❖ B-Trees

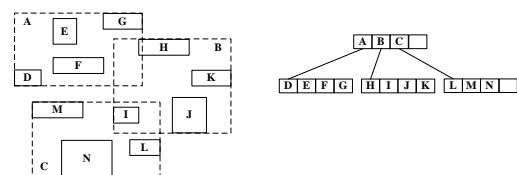
- 键值的一维排序
- 不能搜索多维空间

华中科技大学数字媒体实验室

R-Tree

<http://media.hust.edu.cn>

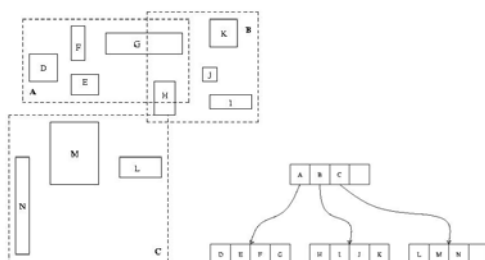
❖ R-Tree是一种典型的向量空间索引结构，它是一种平衡树，其中的每个节点存放着包含数据的最小外接矩形（MBR）



华中科技大学数字媒体实验室

R-Tree (续)

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

R-Tree的缺点

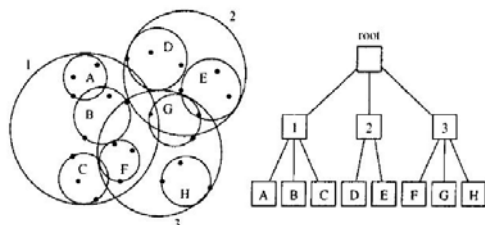
<http://media.hust.edu.cn>

- ❖ R-Tree结构中存放的是数据的MBR，这些数据的MBR可能会相互重叠
- ❖ R-Tree中采用的是最小外接矩形，不能准确地描述距离范围，而在SS-Tree和SR-Tree结构中都采用了球，更好地支持了相似性检索
- ❖ R-Tree利用了数据向量的坐标信息，在距离计算时主要采用欧式距离

华中科技大学数字媒体实验室

SS-Tree

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

边界矩形和边界圆的比较

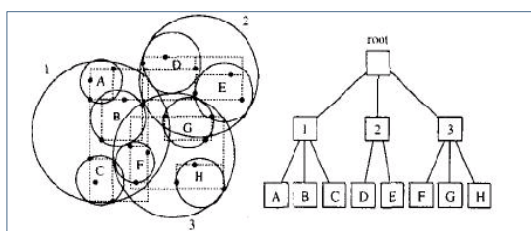
<http://media.hust.edu.cn>

- ❖ 边界矩形的直径（对角线）比边界圆大，SS-Tree将点分到小直径区域，由于区域的直径对最邻近查询性能的影响较大，因此SS-Tree的最邻近查询性能优于R-Tree
- ❖ 边界矩形的平均容积比边界圆小，R*-Tree将点分到小容积区域，由于大的容积会产生较多的覆盖，因此边界矩形在容积方面要优于边界圆

华中科技大学数字媒体实验室

SR-Tree

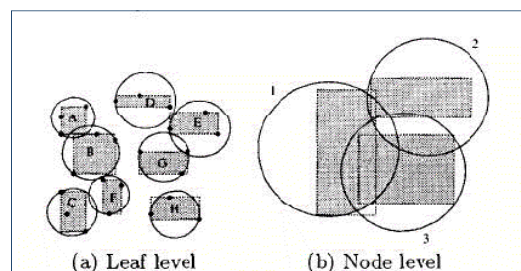
<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

SR-Tree (续)

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

M-Tree

<http://media.hust.edu.cn>

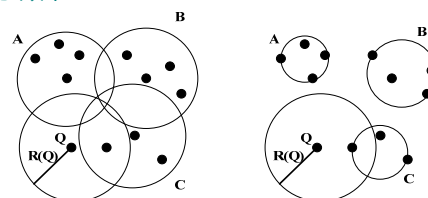
- ❖ M-Tree与R-Tree相同之处
 - 都是平衡树结构
 - 每个节点所存放的项的数目最大不超过N
 - 在动态插入一个数据项时，选择最“相近”的分支向下插入，当节点溢出时，要对节点进行分裂

华中科技大学数字媒体实验室

M-Tree

<http://media.hust.edu.cn>

- ❖ 分裂的子空间，可能出现区域重叠，在检索时降低了效率



- ❖ 当维数超过20时，性能显著下降

华中科技大学数字媒体实验室

M-Tree

<http://media.hust.edu.cn>

❖ M-Tree与R - Tree不同之处

- 只用到了距离信息，没有坐标信息
- 节点中存放的是具有代表性的数据项和该数据项覆盖其所有子数据项所需要的最小半径（最大距离）

华中科技大学数字媒体实验室

VA - File

<http://media.hust.edu.cn>

❖ VA-File (Vector Approximation File)

- ❖ 一种简单但非常有效的方式，它将数据空间划分成 2^b 单元（cell），b表示用户指定的二进制位数，每个单元分配一个位串
- ❖ 位于某个单元内的向量用这个单元近似代替，VA-File本身只是这些近似体的数组
- ❖ 查询时，先扫描VA-File，选择候选向量，再访问向量文件进行验证

华中科技大学数字媒体实验室

VA - File (续)

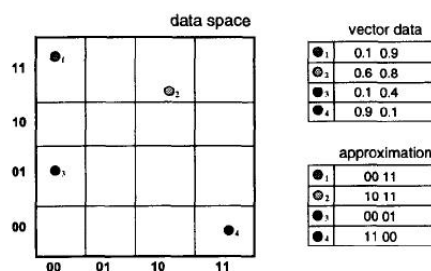
<http://media.hust.edu.cn>

- ❖ VA-File将每一个点数据，将其每一维所在区间对应的比特串串联起来即为其近似值
- ❖ 查询时，VA - File起到过滤的作用，依据比特串所对应区间的边界来确定可能的查询结果集
- ❖ 对结果集中的点数据进行精确的查询

华中科技大学数字媒体实验室

VA - File (续)

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

常见的高维索引结构

<http://media.hust.edu.cn>

- ❖ K-D-B Trees (J T Robinson SIGMOD'1981)
- ❖ R-tree (A. Guttman SIGMOD'1984)
- ❖ R*-tree (T. Sellis VLDB'1987)
- ❖ LSD-Tree (A. Henrich VLDB'1989)
- ❖ R*-Tree (N. Beckmann SIGMOD'1990)
- ❖ TV-Tree (K. I. Lin VLDB'1994)
- ❖ SS-Tree (D. A. White ICDE'1996)
- ❖ VAMSplit R-Tree (D. A. White SPIE'1996)
- ❖ SR-Tree (N. Katayama SIGMOD'1997)
- ❖ M-Tree (P. Ciaccia VLDB'1997)
- ❖ VA-File (R. Weber VLDB'1998)
- ❖ Pyramid-Tree (S. Berchtold SIGMOD'1998)
- ❖ hybrid-Tree (K. Chakrabarti ICDE'1999)
- ❖ A-Tree (Y. Sakurai VLDB'2000)
- ❖ IQ-Tree (S. Berchtold ICDE'2000)

华中科技大学数字媒体实验室

多特征索引结构

<http://media.hust.edu.cn>

❖ 基于样例的视频检索

- 由用户提供检索的样例（图像、视频片段等）
- 由用户选择检索时进行匹配的特征集合（颜色、纹理、形状、运动、音频等）
- 由用户选择特征集合中每个特征的权值

华中科技大学数字媒体实验室

多特征索引结构

http://media.hust.edu.cn

❖多特征索引方式的选择

- 一种特征一个索引：方法相对简单，易于扩充，但检索时需要多次访问多个索引结构，算法复杂度高
- 多种特征一个索引：索引的生成相对较难，不易扩充，索引条件破坏时需要重新生成索引

华中科技大学数字媒体实验室

多特征索引结构

http://media.hust.edu.cn

❖索引结构类型的选择

- M树系列：维度超过20时性能下降
- VA-FILE系列：没有保留层次关系，仅能满足视频检索的需要，不能实现视频浏览
- 保留层次结构用于浏览，树结构的最底层的非叶子节点用于检索

华中科技大学数字媒体实验室

MFI-TREE索引结构

http://media.hust.edu.cn

❖MFI-TREE索引结构

- 集合节点：一个数据对象集合

节点ID	父节点ID	覆盖半径R	集合中心对象ID	浏览对象ID	叶子节点数LeafNum
------	-------	-------	----------	--------	--------------

- 叶子节点：一个数据对象

节点ID	父节点ID	特征 F_1	特征距离值 d_1	特征 F_n	特征距离值 d_n
------	-------	----------	-------------	-------	----------	-------------

华中科技大学数字媒体实验室

MFI-TREE索引结构

http://media.hust.edu.cn

❖对象的相似度距离公式

$$\begin{cases} Dist(P, O) = \sum_{i=1}^n w_i Dist_i(P, O) \\ \sum_{i=1}^n w_i = 1 \end{cases}$$

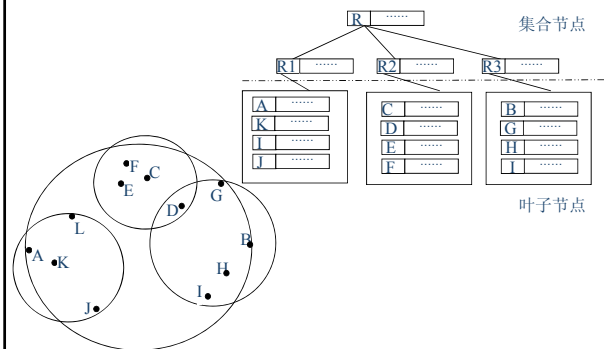
$$Dist_i(P, O) = \frac{Dist'_i(P, O)}{Dist_{i\max}}$$

$$Dist(P, O) = \max(Dist_i(P, O))$$

华中科技大学数字媒体实验室

MFI-TREE索引结构

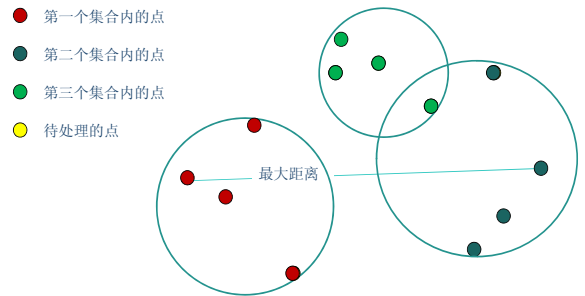
http://media.hust.edu.cn



华中科技大学数字媒体实验室

MFI-TREE分裂算法

http://media.hust.edu.cn



华中科技大学数字媒体实验室

kNN检索算法

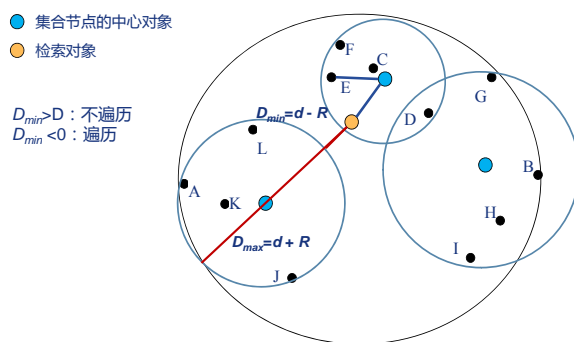
<http://media.hust.edu.cn>

- ❖ kNN检索（k近邻检索）是最常见的一种信息检索方式
- ❖ 在基于MFI-TREE索引结构的kNN检索中，通过快速减小用于过滤的判断值D（排序后第k个检索结果对象与检索对象的最小距离），提高检索的效率

华中科技大学数字媒体实验室

kNN检索算法

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

Outline

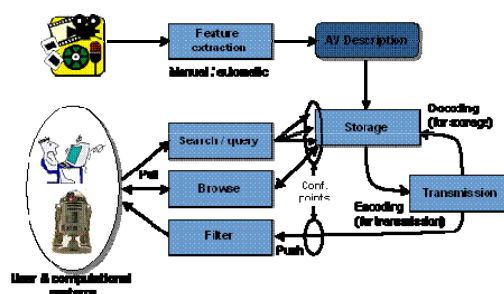
<http://media.hust.edu.cn>



华中科技大學數字媒體實驗室

Video Retrieval

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

视频查询方式分类

<http://media.hust.edu.cn>

- ❖ 确定用户查询的不同类型对视频数据库管理系统的查询功能设计是十分重要的
- ❖ 对于不同的应用领域，用户的查询要求差别较大
- ❖ 但一般来说，视频查询方式的可以按查询内容、匹配精度要求、数据单元大小等进行分类

华中科技大学数字媒体实验室

(1) 按查询层次分类

<http://media.hust.edu.cn>

- ❖ 语义信息查询(Semantic Information Queries , SIQ)
- ❖ 元信息查询(Meta Information Queries , MIQ)
- ❖ 视听查询(Audio Visual Queries , AVQ)

华中科技大学数字媒体实验室

(2) 按提交查询的形式分类

<http://media.hust.edu.cn>

- ❖ 图像查询
- ❖ 运动特征查询
- ❖ 声音查询
- ❖ 文本查询
- ❖ 几何查询

华中科技大学数字媒体实验室

(3) 按查询功能分类

<http://media.hust.edu.cn>

- ❖ 定位查询(Location Queries , LQ)
- ❖ 浏览查询(Browsing Queries , BQ)
- ❖ 跟踪查询(Tracking Queries , TQ)

华中科技大学数字媒体实验室

(4) 按查询的数据单元大小分类

<http://media.hust.edu.cn>

- ❖ 基于帧的查询(Frame-based Queries , FBQ)
- ❖ 基于视频片段的查询(Clip-based Queries , CBQ)
- ❖ 基于视频流整体的查询(Video Stream-based Queries , VBQ)

华中科技大学数字媒体实验室

(5) 按查询的表述方式分类

<http://media.hust.edu.cn>

- ❖ 直接查询(Direct Queries , DQ)
- ❖ 示例查询(Queries By Example , QBE)
- ❖ 渐进查询(Iterative Queries , IQ)

华中科技大学数字媒体实验室

(6) 按查询匹配精度分类

<http://media.hust.edu.cn>

- ❖ 精确匹配查询(Exact Match Queries , EMQ)
- ❖ 相似匹配查询(Similarity Match Queries , SMQ)
 - 范围查询 (Range Query, RQ)
 - K - 近邻查询 (k Nearest - Neighbor Query , kNNQ)

华中科技大学数字媒体实验室

范围查询

<http://media.hust.edu.cn>

- ❖ 对于数据库DB, 给定检索数据q与一个相似度距离值T, 以及事先在数据库中定义的一个值域大于零的相似度距离函数
- ❖ 检索时根据相似度距离函数计算数据库中对象与检索数据的距离值, 找出其中所有距离值小于最大相似度距离值的对象

$$RQ(q, T) = \{o \in DB \mid \|o - q\| \leq T\}$$

华中科技大学数字媒体实验室

k - 近邻查询(kNN)

<http://media.hust.edu.cn>

- ❖ 对于数据库DB，给定检索数据q与一个正整数k，以及定义的相似度距离函数
- ❖ 检索时计算数据库中对象与检索数据的距离值，找出其中k个距离值最小的对象

$$kNNQ(q, k) = \{o_0 \dots o_{k-1} \in DB \mid \forall e \in DB \setminus \{o_0 \dots o_{k-1}\}, \|o_i - q\| \leq \|e - q\|, 0 \leq i \leq k-1\}$$

华中科技大学数字媒体实验室

相似匹配查询的一般方法

<http://media.hust.edu.cn>

- ❖ Bound-and-Branch
 - 主要应用在树形索引结构中
 - **基本思想**：是从根节点开始，计算每个节点与查询数据之间的距离，如发现不可能包含查询结果的节点，则放弃查询节点下的路径，从而减少磁盘访问代价

华中科技大学数字媒体实验室

相似匹配查询的一般方法（续）

<http://media.hust.edu.cn>

- ❖ 基于过滤的方法
 - 适用于一些以顺序方式存储近似矢量的结构
 - 过滤阶段
 - 精确计算

华中科技大学数字媒体实验室

视频数据查询和检索的过程

<http://media.hust.edu.cn>

- ❖ 视频数据查询和检索是一个逐步求精的过程，检索经历了一个特征调整、重新匹配的循环过程：
 - **初始查询** 用户查找一个对象时，最初可以用示例查询或查询语言来形成一个查询，系统提取该示例的特征或把查询描述映射为具体的特征矢量
 - **相似性匹配** 将查询特征与特征库中的特征按照一定的匹配算法进行相似匹配

华中科技大学数字媒体实验室

视频数据查询和检索的过程(续)

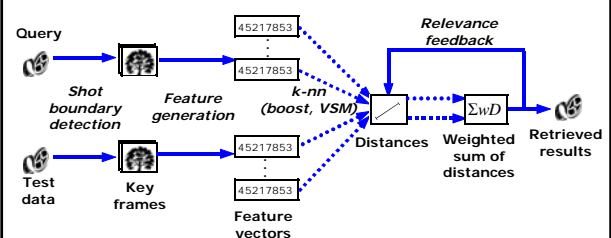
<http://media.hust.edu.cn>

- ❖ 满足一定相似性条件的一组候选结果，按相似度大小排列后返回给用户
- ❖ **特征调整** 对系统返回的查询结果，用户可以通过遍历（浏览）来挑选，直至得到满意的结果，或者从候选结果中选择一个示例，经过特征调整后，形成一个新的查询
- ❖ 如此逐步缩小查询的范围，直到用户对查询结果满意为止

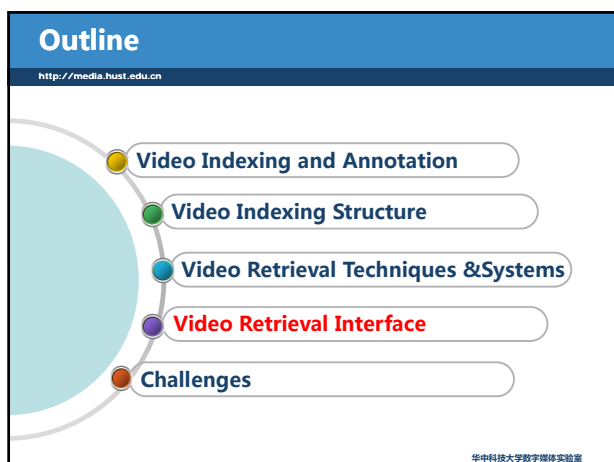
华中科技大学数字媒体实验室

视频数据查询和检索的过程(续)

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室



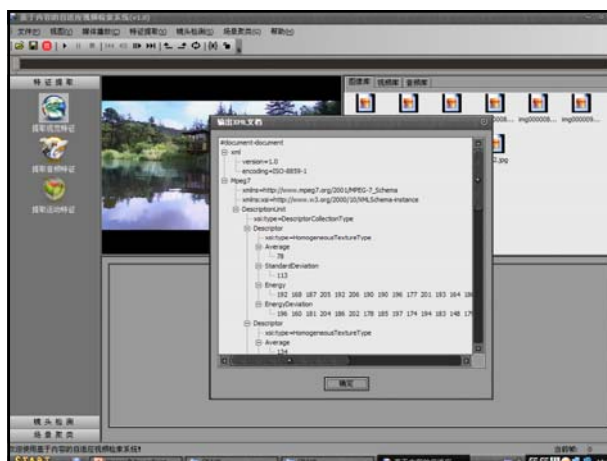
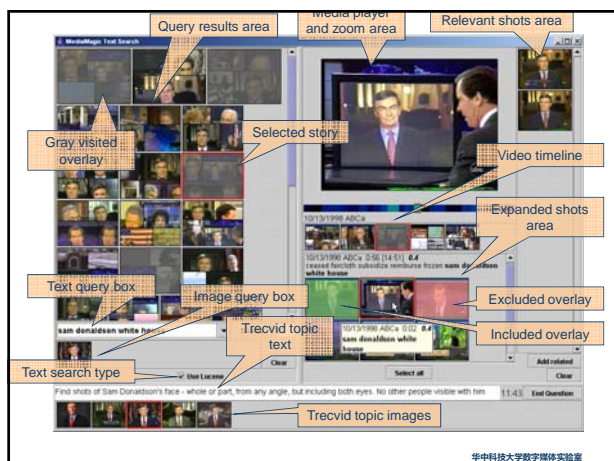
用户界面和浏览显示

http://media.hust.edu.cn

❖ 视频检索系统的显示应提供和保证如下主要功能：

- 操作控制
- 视频片段显示
- 交互和反馈

华中科技大学数字媒体实验室



Outline

<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

Opportunities

<http://media.hust.edu.cn>

- ❖ Billions of Images
- ❖ Millions of Videos
- ❖ Hundreds of Millions of users
- ❖ Reasonable Precision rates

华中科技大学数字媒体实验室

Challenges

<http://media.hust.edu.cn>

- ❖ Representation
 - Models of visual features
 - Image/video object representations
 - Description of spatio-temporal relationships video objects
 - High-level event representations
- ❖ Analysis:
 - Semantic Segmentation of video
 - Adaptive grouping of features & object

华中科技大学数字媒体实验室

Challenges

<http://media.hust.edu.cn>

- ❖ Indexing
 - Efficient indexing algorithms for high-dimensional feature space
 - Robust, scalable indexing algorithms for spatio-temporal queries
- ❖ Retrieval
 - Efficient query interface for query manipulation and relevance feedback
 - User-friendly navigation
 - Query optimization considering multiple criteria (storage, speed, bandwidth)

华中科技大学数字媒体实验室

Challenges

<http://media.hust.edu.cn>

- ❖ Classification
 - Domain-specific or domain-independent video classification
 - Semantic decoding and model selection through learning
- ❖ Summarization
 - Automated summarization of visual content
 - Visualization of content at different levels
- ❖ Practical problem
 - Mismatch between percepts and concepts

华中科技大学数字媒体实验室

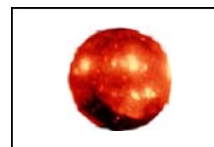
Perceptual/Conceptual Issue

<http://media.hust.edu.cn>

Similar Percepts / Dissimilar Concepts



Clown Nose



Red Sun

华中科技大学数字媒体实验室

Perceptual/Conceptual Issue

<http://media.hust.edu.cn>

Dissimilar Percepts / Similar Concepts



John's Car



Mike's Car

华中科技大学数字媒体实验室

结论

<http://media.hust.edu.cn>

如何有效地填补“语义鸿沟”，
从而提高语义级视频检索的效率和服务
质量已经成为基于内容视频检索领域的
研究难点和最为关键的问题！

华中科技大学数字媒体实验室

