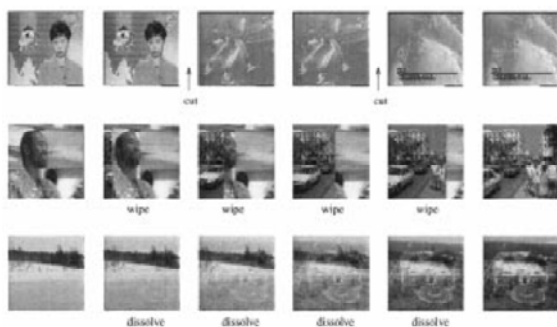


表1 常见的10种镜头变换

镜头变换	效果描述
切换 (Cut)	前一镜头的尾帧被下一镜头的首帧快速代替
淡入/淡出 (Fade in/Fade out)	某镜头的尾帧缓慢而均匀地变黑直至全部消失 (淡出), 或某镜头的首帧缓慢而均匀地从全黑屏幕中出现 (淡入)
隐现 (Dissolve)	前一镜头的尾帧缓慢而均匀地变成下一镜头的首帧
翻页 (Page Translate)	前一镜头的尾帧从屏幕一边拉出, 并逐渐显露出下一镜头的首帧
拉进 (Slide)	下一镜头的首帧从屏幕一边或一角平稳地拉进, 同时前一镜头的尾帧从另一边或一角拉出
滑入 (Wipe)	下一镜头的首帧逐渐穿过并覆盖前一镜头的尾帧
弹出/弹出 (Pop on/Pop off)	某镜头的首帧立即从屏幕中出现或某镜头的尾帧立即从屏幕中消失
上拉/下拉 (Pull Up/Pull Down)	下一镜头的首帧如窗帘一样从屏幕顶部拉下, 遮住前一镜头的尾帧
翻转 (Flip)	前一镜头的尾帧翻转, 在另一面显示下一镜头的首帧
旋转 (Spin)	下一镜头的首帧以旋转方式出现并覆盖前一镜头的尾帧

例子

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



华中科技大学数字媒体实验室

镜头变换检测方法分类

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

❖ 目前镜头变换检测方法可以分为三大类：

- 基于解压的全图像序列的检测方法
- 直接基于压缩视频的检测方法
- 基于确定变换模型的检测方法

华中科技大学数字媒体实验室

(一) 切变的检测

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

- ❖ **切变**：当前图像被下一幅图像快速代替，人的视觉上可以感觉到一个突然的变化
- ❖ 同样在帧间差别比较上会出现一个尖峰(Peak)
- ❖ **基本原理**：寻找较好的帧间差别比较方法来检出尖峰，同时减少与其它镜头变换方式的误识别

华中科技大学数字媒体实验室

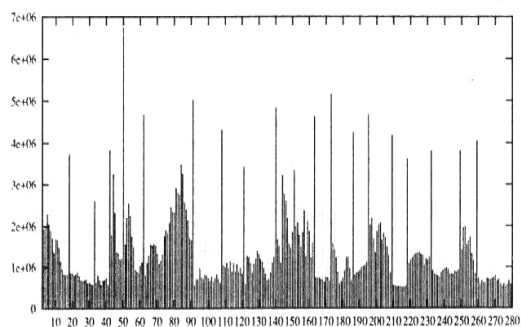
(1) 模板匹配法

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

- ❖ 对两幅图像相同位置的像素灰度值或颜色值进行比较，并计算各灰度差或颜色差的总和
- ❖ 计算公式：

$$d(I_i, I_j) = \sum_{x=0, y=0}^{x=N_x, y=N_y} |I_i(x, y) - I_j(x, y)|$$

华中科技大学数字媒体实验室



模板匹配法检测结果

华中科技大学数字媒体实验室

模板匹配法的特点：

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

- ❖ 计算方法与像素的位置密切相关
- ❖ 图像差别的计算对噪声和物体运动十分敏感
- ❖ 在镜头变换识别中，如果单纯采用模板匹配的方法，容易造成误识别

华中科技大学数字媒体实验室

(2) 颜色直方图匹配

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

- ❖ 若两幅图像 I_i 和 I_j 的直方图为 H_i 和 H_j ，则颜色直方图匹配的计算公式如下：

$$d(I_i, I_j) = \sum_{k=1}^n |H_i(k) - H_j(k)|$$

- ❖ 对于RGB图像，颜色由不同亮度的R(红)、G(绿)、B(蓝)三基色组成，因此可以改写成：

$$d_{RGB}(I_i, I_j) = \sum_{k=1}^n (|H_i^r(k) - H_j^r(k)| + |H_i^g(k) - H_j^g(k)| + |H_i^b(k) - H_j^b(k)|)$$

华中科技大学数字媒体实验室

颜色直方图匹配的特点

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

- ❖ 由于直方图丢失了颜色的位置信息，因此两幅图像可能内容完全不同但直方图相似
- ❖ 仅用简单的颜色直方图匹配也容易造成误识别
- ❖ 一种改进的方法是将图像划分成若干子块，分别对各子块进行匹配
- ❖ 子块划分与匹配的方法可以对物体运动、摄像机运动、镜头缩放等情况有更好的适应性

华中科技大学数字媒体实验室

(3) χ^2 直方图匹配

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

- ❖ χ^2 直方图匹配的计算公式：

$$d(I_i, I_j) = \sum_{k=1}^n \frac{(H_i(k) - H_j(k))^2}{H_j(k)}$$

- ❖ 对RGB图像， χ^2 直方图匹配公式可以变为：

$$d(I_i, I_j) = \sum_{k=1}^n \left(\frac{(H_i^r(k) - H_j^r(k))^2}{H_j^r(k)} + \frac{(H_i^g(k) - H_j^g(k))^2}{H_j^g(k)} + \frac{(H_i^b(k) - H_j^b(k))^2}{H_j^b(k)} \right)$$

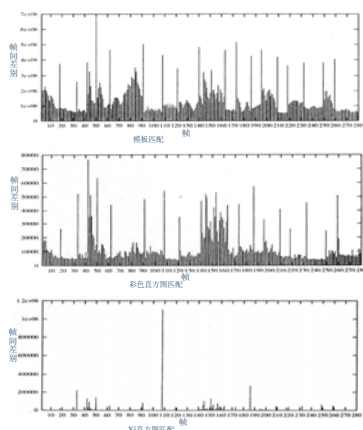
华中科技大学数字媒体实验室

特点

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

- ❖ 若干研究表明， χ^2 直方图匹配与其它方法如模板匹配或颜色直方图匹配相比具有更好的识别率
- ❖ 在识别镜头切换上效果良好
- ❖ **需要注意**，上述的帧间差别计算方法在镜头切换识别中往往并不是单独采用，而是两种甚至多种方法混合使用

华中科技大学数字媒体实验室



华中科技大学数字媒体实验室

启发与挑战：

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

- ❖ 将直方图特征更换为其它特征，看效果如何？
- ❖ 阈值如何确定，能否自动设置阈值，阈值能否自适应？

华中科技大学数字媒体实验室

(二) 渐变的检测

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

- ❖ 渐变的检测比切变的检测要复杂得多
- ❖ 这主要是由于渐变在帧间差别上并没有一个可检测的尖峰存在，而且特别容易与物体或摄像机运动相混淆
- ❖ 渐变的检测往往需要对多帧进行帧间差别计算并找到它们的变化规律才能实现

华中科技大学数字媒体实验室

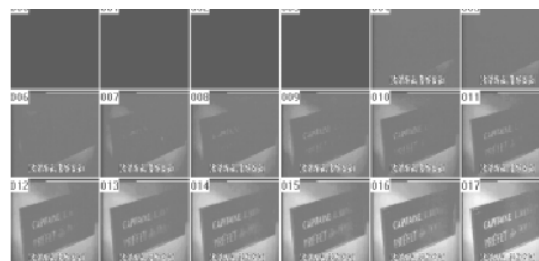


图 11.3.1 渐变的一个示例

华中科技大学数字媒体实验室

(1) 双重比较法-Twin Comparision

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

- ❖ 所谓双重比较法，简言之就是采用两个阈值来检测镜头变换
- ❖ 设置两个阈值 T_b 和 T_s ($T_s < T_b$)，分别是镜头切换和渐变的阈值
- ❖ 首先用较低的阈值 T_s 来确定渐变过程的起始帧

华中科技大学数字媒体实验室

双重比较法

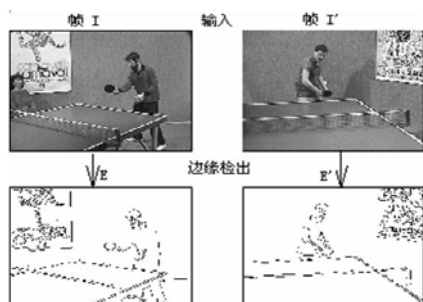
<http://media.hust.edu.cn>

- ❖ 如果两连续帧的直方图差 $d(H_i, H_{i+1})$ 满足 $T_s < d(H_i, H_{i+1}) < T_b$ ，则它们被认为很可能是渐变的开始
- ❖ 一旦检出这种帧，就开始进行帧间直方图差累计 $A_c(i)$
- ❖ 直至满足 $A_c(i) > T_b$ 且 $d(H_i, H_{i+1}) < T_s$ ，则认为是渐变的结束

华中科技大学数字媒体实验室

(2) 边缘变换识别方法

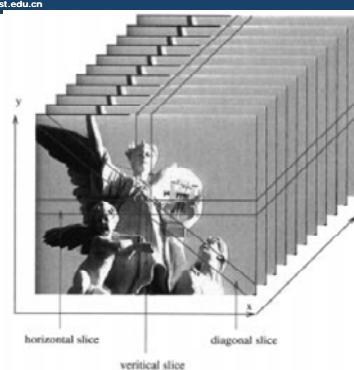
<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

(3) 基于时空切片检测镜头变换

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



华中科技大学数字媒体实验室

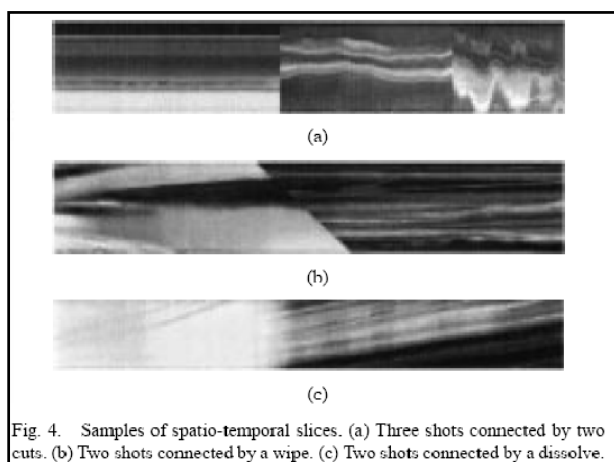
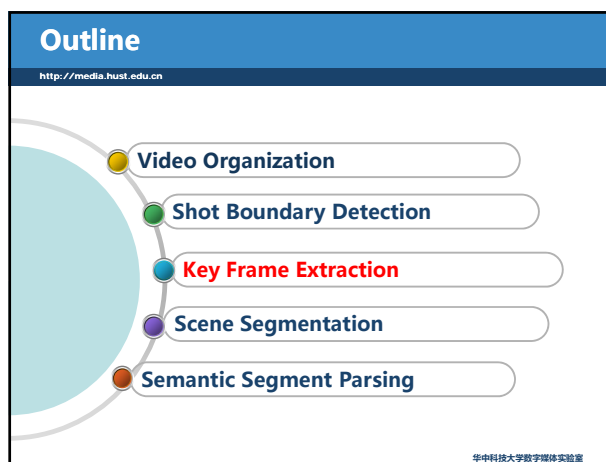


Fig. 4. Samples of spatio-temporal slices. (a) Three shots connected by two cuts. (b) Two shots connected by a wipe. (c) Two shots connected by a dissolve.



关键帧(Keyframe)的定义

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

- ❖ **关键帧** (有时也称代表帧) 是用于描述一个镜头的关键图像帧, 它通常会反映一个镜头的主要内容
- ❖ 依据镜头内容的复杂程度, 可以从一个镜头中提取一个或多个关键帧

华中科技大学数字媒体实验室

关键帧的用途

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

- ❖ 关键帧的使用大大减少了视频索引的数据量
- ❖ 同时也为查询和检索视频提供了一个组织框架
- ❖ 通过关键帧还可以为一段视频建立视频摘要, 允许用户通过观看几个有限的关键帧来快速浏览整段视频

华中科技大学数字媒体实验室

提取关键帧的目的

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

- ❖ 其一, 希望用它来静态表示视频节目的主题和部分内容, 而不是动态的细节
- ❖ 其二, 则是希望从关键帧中提取颜色、纹理和形状特征, 以作为视频数据库索引的数据源, 而不需要对每个画面都重复提取
- ❖ 由此可见, 关键帧应具有**代表性**

华中科技大学数字媒体实验室

关键帧的选取原则

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

- ❖ 一般采用保守原则来提取关键帧, 即关键帧的提取 **"宁愿错, 勿能少"**
- ❖ 在代表特征不具体的情况下, 一般以去掉重复 (或冗余) 画面为原则
- ❖ 基于这一基本原则, 不同的提取算法可以选取不同的原则, 建立适合自身情况的判定标准

华中科技大学数字媒体实验室

关键帧提取的理论基础

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

- ❖ 原则上讲，关键帧应能提供一个镜头的全面概要，或者说应能提供一个内容尽量丰富的概要
- ❖ 根据信息论的观点，不同（或相关性很小）的帧图像比类似的帧图像携带更多的信息
- ❖ 因此，当需要多幅关键帧时，用于关键帧提取的准则主要是考虑它们之间的不相似性

华中科技大学数字媒体实验室

典型的关键帧提取算法

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

- ❖ 基于镜头边界提取关键帧
- ❖ 基于图像信息提取关键帧
- ❖ 基于运动分析提取关键帧
- ❖ 像素平均法和直方图平均法
- ❖ 宏块互异法
- ❖ 基于文字和图象信息的方法

华中科技大学数字媒体实验室

参考文献：

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

- [1] H. Zhang, J. Wu, D. Zhong, and S. W. Smoliar, "An integrated system for content-based video retrieval and browsing," Pattern Recognition, vol. 30, no. 4, pp. 643-658, 1997.
- [2] W. Wolf, Key frame Selection by motion analysis, in Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc., 1996.c.
- [3] Zhang Y J, Lu H B. 1999. Hierarchical video organization based on compact representation of video units. Proc. Workshop on very Low Bitrate Video' 99, 67-70
- [4] Bilge Gunsul, Tekalp A. Murat. Content-based access to video object: Temporal segmentation, visual summarization, and feature extraction. Signal Processing, 66:261~280, 1998.
- [5] 杨胜, 钟玉琢. 一种从MPEG压缩视频流中提取关键帧的方法, 中国图形图象学报, 2001, 3: 254~258

华中科技大学数字媒体实验室

（一）基于镜头边界提取关键帧

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

- ❖ 一段视频分割成镜头后，将每个镜头的首帧（或首帧与末帧）作为镜头的关键帧
- ❖ **优点：**实现起来较为简单，无论镜头的内容如何，关键帧的数量都是一定的（1帧或2帧）
- ❖ **缺点：**效果不是很稳定，因为每个镜头的首帧或末帧不一定总是能够反映镜头的主要内容

华中科技大学数字媒体实验室

（二）基于图像信息提取关键帧

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

- ❖ **原理：**基于每一帧的颜色、纹理等视觉信息的改变来提取关键帧，当这些信息有显著变化时，当前的帧即可作为关键帧
- ❖ 首先把镜头的第一帧作为关键帧，并作为参考帧
- ❖ 然后计算前一个关键帧与剩余帧之差，如果差值大于某一阈值，则再选取一个关键帧
- ❖ 把刚选出的帧作为参考帧，依次类推

华中科技大学数字媒体实验室

特点：

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

- ❖ **优点：**可以根据镜头内容的变化程度选择相应数目的关键帧
- ❖ **缺点：**所选取的帧不一定具有代表意义，而且在有镜头运动时，容易选取过多的关键帧

华中科技大学数字媒体实验室

(三) 基于运动分析提取关键帧

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

- ❖ 通过光流分析来计算镜头中的运动量，在运动量取局部最小值处选取关键帧
- ❖ 它反映了视频数据中的静止，视频中通过摄像机在一个新的位置上停留或通过人物的某一动作的短暂停留来强调其本身的重要性

华中科技大学数字媒体实验室

特点：

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

- ❖ **优点：**相对于前两种方法，该方法提取的关键帧代表性较强
- ❖ **缺点：**在分析运动时，需要的计算量较大，而且局部最小值也不一定准确

华中科技大学数字媒体实验室

(四) 基于文字和图象信息提取关键

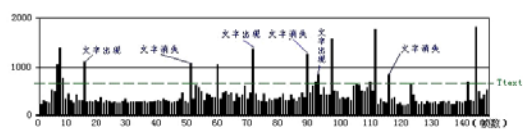
<http://media.hust.edu.cn>

- ❖ 在新闻视频中，有文字出现的帧往往是一条新闻、一个新闻视频镜头的关键帧
- ❖ 这些视频帧最大限度地反映了新闻的主要内容，具有极强的代表性
- ❖ 该方法首先要能够检测出有文字出现的帧，尤其是在固定区域有文字出现的帧

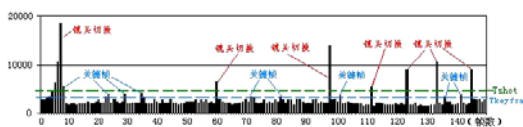
华中科技大学数字媒体实验室



华中科技大学数字媒体实验室



(a) “局部”欧氏直方差



(b) 全帧欧氏直方差

华中科技大学数字媒体实验室

Outline

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

Video Organization

Shot Boundary Detection

Key Frame Extraction

Scene Segmentation

Semantic Segment Parsing

华中科技大学数字媒体实验室

何谓“视频聚类”？

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

- ❖ **视频聚类**是把相似的原始视频数据按照确定的相似性测量准则进行聚集，形成更高级的视频语义层次，以便于对视频数据进行组织、浏览、建立索引、查询和摘要等
- ❖ 视频聚类是基于内容视频检索的基础，是迈向视频理解的必经之路

华中科技大学数字媒体实验室

聚类的基本概念

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

- ❖ **聚类 (Cluster)**就是将数据对象分成为多个类或簇，在同一个簇中的对象具有较高的相似度，而不同簇中的对象差别很大
- ❖ **聚类分析方法**是一种无监督的学习方法，它是将一些未知模式分成若干类
- ❖ 根据特征向量间的距离在一定的误差范围内相等，则认为它们是同一类

华中科技大学数字媒体实验室

聚类方法分类

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

- ❖ 划分方法 (Partitioning Method)
- ❖ 层次方法 (Hierarchical Method)
- ❖ 基于密度的方法 (Density-based Method)
- ❖ 基于网络的方法 (Grid-based Method)
- ❖ 基于模型的方法 (Model-based Method)

华中科技大学数字媒体实验室

常用的聚类方法

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

- ❖ **K-均值法**
Telalp A M. 1995. Digital Video Processing. Prentice-Hall.
- ❖ **ISODATA法**
田村秀行 (日) 著, 赫荣威等编译. 1988. 计算机图象处理技术. 北京: 北京师范大学出版社

华中科技大学数字媒体实验室

❖ 松弛迭代法

Kittler J, Illingworth J. Relaxation labeling algorithms-a review. Image and vision Computing, 1985, 3(4): 206~216

❖ 基于关联规则的分类法

Liu B, Hsu W, Ma Y M. Integrating classification and association rule mining. Pro, 4ICKDDM, 1998, 80~86

❖ 基于模糊图论聚类法

Jafrkhani H, Tarokh V. Image clustering using fuzzy graph theory. SPIE, 2000, 3972: 245~252

华中科技大学数字媒体实验室

K-均值法

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

- ❖ **K-均值法**是一种划分方法
- ❖ **基本原理**: 给定一个n个对象的数据库，构建数据的k个划分，每个划分表示一个聚簇，并且 $k \leq n$
- ❖ 也就是说，它将数据划分为k个组，同时满足如下的要求：每个组至少包含一个对象
- ❖ 每个对象必须属于且只属于一个组

华中科技大学数字媒体实验室

- ❖ K-均值算法以 k 为参数,把 n 个对象分为 k 个簇,以使簇内具有较高的相似度,而簇间的相似度较低
- ❖ 首先,随机地选择 k 个对象,每个对象初始地代表了一个簇的平均值或中;
- ❖ 对剩余的每个对象,根据其与其各个簇中心的距离,将它赋于最近的簇
- ❖ 然后重新计算每个簇的平均值,将每个对象重新赋给最近的簇
- ❖ 这个过程不断重复,直到平均值不再发生变化

华中科技大学数字媒体实验室

K-均值法的特点

http://media.hust.edu.cn

- ❖ K-均值算法要求用户必须事先给出 k 值(要求生成的簇的数目),可能不适于某些应用
- ❖ 对于大多数视频数据的聚类,用户很难在聚类之前给出生成的簇的数目
- ❖ 因此,视频聚类中很少采用 k -平均算法,而大多采用层次聚类算法

华中科技大学数字媒体实验室

层次聚类法

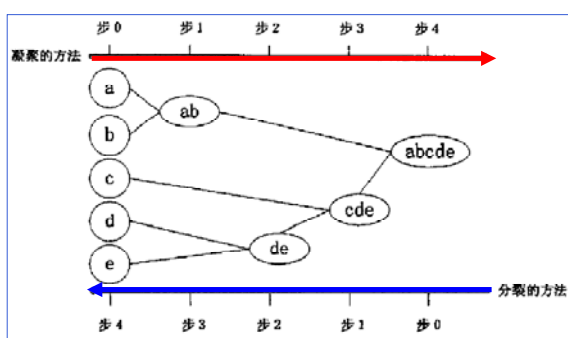
http://media.hust.edu.cn

- ❖ 层次方法对给定数据对象集合进行层次的分解
- ❖ 根据层次的分解如何形成,分为凝聚的和分裂的
- ❖ **凝聚的层次聚类**:自底向上,首先将每个对象作为一个簇,然后合并这些原子簇为越来越大的簇,直到所有的对象都在一个簇中,或者满足某个终止条件

华中科技大学数字媒体实验室

- ❖ **分裂的层次聚类**:自顶向下,与凝聚的层次聚类相反
- ❖ 它首先将所有对象置于一个簇中,然后逐渐细分为越来越小的簇,直到每个对象自成一簇,或者达到某个终止条件
- ❖ 例如达到了某个希望的数目,或者两个最近的簇之间的距离超过了某个阈值

华中科技大学数字媒体实验室



在数据对象集合上的凝聚和分裂层次聚类

华中科技大学数字媒体实验室

镜头聚类

http://media.hust.edu.cn

- ❖ 场景中的镜头具有相同或相似的内容,如动作、场所和时间
- ❖ 如果镜头具有相同或相似的内容,就认为这两个镜头是语义相关的
- ❖ 在镜头聚类为场景的过程中,主要的任务是使相同或相似的镜头形成场景
- ❖ 因此,关键的问题是如何度量镜头的相似度

华中科技大学数字媒体实验室

(一) 镜头相似度的测量方法

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

❖ 镜头相似度测量技术主要可以分为以下两种：

- 基于视频帧序列的镜头相似度测量
- 基于镜头关键帧的镜头相似度测量

华中科技大学数字媒体实验室

基于视频帧序列的镜头相似度测量

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

- ❖ 计算镜头中帧对的相似度
- ❖ 设计一个对准函数和一个构造函数计算镜头相似度
- ❖ 特点：
 - 需要计算镜头视频帧中每一个帧对的相似度
 - 计算量比较大
 - 设计一个好的对准函数和构造函数也较为困难

华中科技大学数字媒体实验室

参考文献

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

- [1] J.M.Corrioni, A.D.Bimbo, **Structured representation of videos: A new approach**, In Proc. Of the SPIE, 2419: 38~47, 1995.
- [2] Sen-ching S. Cheung and Avidesh Zakhor, **Efficient Video Similarity Measurement with Video Signature**, International Conference on Image Processing, Rochester, New York, September 22-25, 2002
- [3] Sen-ching S. Cheung and Avidesh Zakhor, **Efficient Video Similarity Measurement and Search**, in Proceedings of the 7th IEEE International Conference on Image Processing, 85~88, Vancouver, British Columbia, 2000
- [4] Yap-Peng Tan Sanjeev R. Kulkarni Peter J. Ramadge, **A Framework for Measuring Video Similarity and It's Application to Video Query by Example**, Pro. IEEE International Conference on Image Processing, Kobe, Japan, 1999.

华中科技大学数字媒体实验室

基于镜头关键帧的镜头相似度测量

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

- ❖ 使用镜头关键帧之间的相似度来度量镜头之间的相似度
- ❖ 特点：
 - 只计算镜头关键帧对之间的相似度
 - 计算量很小，计算速度快
- ❖ 使用代表镜头主要内容的关键帧计算镜头之间的相似度也是合理的

华中科技大学数字媒体实验室

参考文献

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

- [1] N. Dimitrova and M. Abdel-Mottaleb, **Content-based Video Retrieval by Example Video Clip**, Proc. SPIE Vol. 3022, Storage and Retrieval for Image and Video Databases, 59~70, 1997.
- [2] Arman, F., Depommier, R., HSU, A., and M. -Y. Chiu. **Content-based Browsing of Video Sequences**. In Proceedings of ACM International Conference on Multimedia'94, 1994.
- [3] H. Aoki, S. Shimotsuji, and O. Hori. **A shot Classification method of Selecting effective key-frames for video browsing**. ACM Multimedia 96, 1~10, 1996.
- [4] D. Zhong, H.J. Zhang, and S.-F. Chang. **Clustering methods for video Browsing and Annotation**, in Proc. Storage and Retrieval for Image and Video Databases, SPIE, 2670: 239~246, 1996.
- [5] Aya Aner and John R. Kendr, **Video Summaries through Mosaic-Based Shot and Scene Clustering**, In Proceedings European Conference on Computer Vision, Denmark, May 2002.

华中科技大学数字媒体实验室

(二) 基于镜头相似度的镜头聚类算法

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

- ❖ 对于大多数视频数据的聚类，用户很难在聚类之前给出簇的数目，因此视频聚类大多使用层次化聚类算法
- ❖ 视频镜头的层次化聚类可以分为：
 - 不考虑时间因素
 - 考虑时间因素

华中科技大学数字媒体实验室

不考虑时间因素的视频聚类

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

- ❖ 设经过分段形成的镜头为： $\{S_1, S_2, \dots, S_i, \dots, S_n\}$ ，其中 n 为镜头的个数
- ❖ 步骤1：令 $C_1=\{S_1\}, C_2=\{S_2\}, \dots, C_i=\{S_i\}, \dots, C_n=\{S_n\}$;
- ❖ 步骤2：令 $i=1$ ，
- ❖ 步骤3：令 $j=1$;
- ❖ 步骤4：若 $i \neq j$ ，计算簇 C_i 与 C_j 之间的距离：

$$D(C_i, C_j) = \text{Min}(S_k, S_l)$$
 其中， S_k 为 C_i 中的所有元素
 S_l 为 C_j 中的所有元素

华中科技大学数字媒体实验室

不考虑时间因素的视频聚类（续）

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

- ❖ 步骤5：若 $D(C_i, C_j) < T$ ，则 $C_i = \{C_i, C_j\}$ ， $n=n-1$ ，返回第2步，
 若 $D(C_i, C_j) > T$ ，且 $j < n$ ，则 $j=j+1$ ，返回第4步；
- ❖ 步骤6：若 $i < n$ ，返回第2步；
- ❖ 步骤7：对所有产生的簇进行标记处理，以便于浏览和检索；
- ❖ 步骤8：算法结束。

华中科技大学数字媒体实验室

考虑时间因素的视频聚类

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

- ❖ 设经过分段形成的镜头为： $\{S_1, S_2, \dots, S_i, \dots, S_n\}$ ，其中 n 为镜头的个数
- ❖ 步骤1：设当前镜头 $S=S_i$ ，当前簇 $C=\{S_1\}$ ， $i=0$
- ❖ 步骤2：计算当前镜头的后续镜头 S_j 与当前簇中所有镜头的距离的最小值：

$$D(S_j, C) = \text{Min}(S_j, S_l)$$
 其中， S_l 为 C 中的所有元素

华中科技大学数字媒体实验室

考虑时间因素的视频聚类（续）

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

- ❖ 步骤3：若 $D(S_j, C) < T$ ， $C=\{S_j, C\}$ ， $n=n-1$ ，返回第2步
 若 $D(S_j, C) > T$ ， $i=i+1$ ，标记当前簇为一个场景
 $E_i=C$ ， $C=S_j$ ，返回第2步
- ❖ 步骤4：所有聚类结果为 E_i ，其中 $i=1, 2, \dots, m$ ， m 为产生的场景的个数
- ❖ 步骤5：算法结束

华中科技大学数字媒体实验室

Outline

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

Video Organization

Shot Boundary Detection

Key Frame Extraction

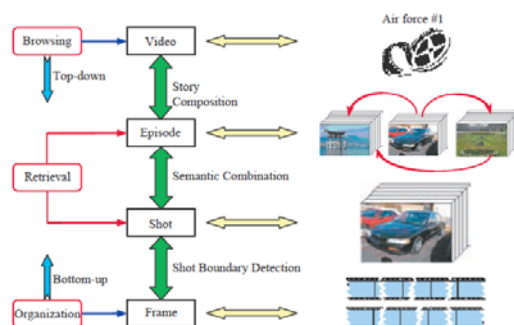
Scene Segmentation

Semantic Segment Parsing

华中科技大学数字媒体实验室

Semantic Segment Parsing

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



华中科技大学数字媒体实验室

情节与逻辑故事单元

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

- ❖ 从编辑的角度看，**镜头**是视频沿时间的一个技术单元
- ❖ **场景**（或事件）是视频中的最小语义单元，它是将一个或几个相关的镜头组合起来得到的
- ❖ 一个场景可以由一个或多个场景片段构成，而一个事件片段可由一个或多个镜头构成
- ❖ **情节**：包含一段时间上相关的故事内容，一个情节可由一个或多个场景组成

华中科技大学数字媒体实验室

情节与逻辑故事单元（续）

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

- ❖ 一个逻辑故事单元与一个情节有许多接近之处
- ❖ **逻辑故事单元**（Logic Story Unit, LSU）：一系列时间上连续的镜头，其中具有类似视觉内容的元素有重复的连接
- ❖ 一个逻辑故事单元与一个情节可能完全对应，但有时又有所不同

华中科技大学数字媒体实验室

情节与逻辑故事单元（续）

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

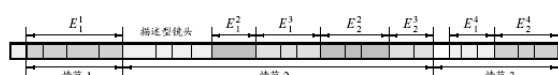


图 12.3.1 情节、事件、镜头示意图

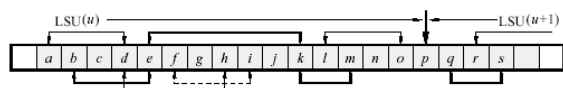
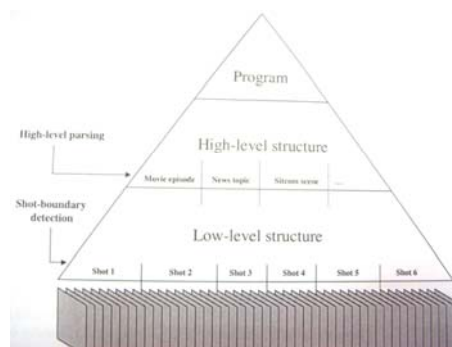


图 12.3.3 根据镜头特点确定逻辑故事单元的边界

华中科技大学数字媒体实验室

Video Content Structure Pyramid

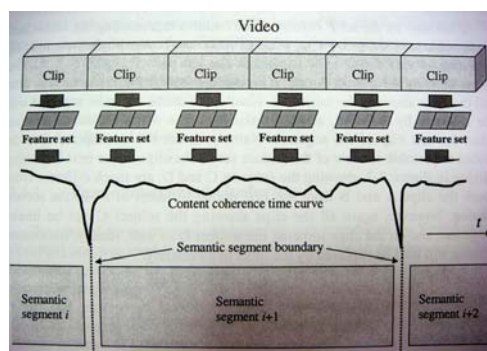
<http://media.hust.edu.cn>



华中科技大学数字媒体实验室

High-level Video Parsing Process

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



华中科技大学数字媒体实验室

News Example

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



华中科技大学数字媒体实验室

参考文献

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

- [1] Hanjalic A, Lagendijk R L, Biemond J. **Automated Segmentation of Movies into Logic Story Units**, Proceeding of Visual Information System'99, 1999, 229~236
- [2] Hanjalic A, et al. **Image and Video Database: Restoration, Watermarking and Retrieval**. Elsevier, 2000.

华中科技大学数字媒体实验室

场景转换图

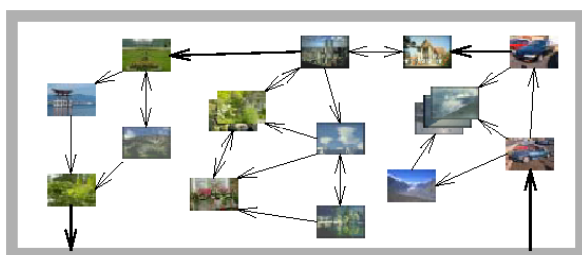
<http://media.hust.edu.cn>

- ❖ **场景转换图** (Scene Transition Graph) 是一种表达故事单元 (Story Unit) 的紧凑方式
- ❖ STG既显示了视频的视觉内容也表达了视频的时间流程
- ❖ 在STG中, 每个节点是相似镜头的集合, 节点间的弧指示了镜头间的联系和故事的流程

华中科技大学数字媒体实验室

场景转换图

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



场景转换示意图

华中科技大学数字媒体实验室

参考文献

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

- [1] Yeung M M, Yeo B L, Wolf W, et al. **Video Browsing using Clustering and Scene Transitions on Compressed Sequences**, SPIE, 1995, 3312, 2417: 375~380
- [2] Yeo B. L. et al. **Classification, Simplification, and Dynamic Visualization of Scene Transition Graph for video browsing**. SPIE, 1998, 3312: 60~72
- [3] Yeung M M, Yeo B L, **Time-constrained Clustering for Segmentation of Video into Story Units**. Proc. ICPR, 375~380
- [4] Yeo B.L, Yeung M M. **Retrieving and Visualizing Video**. Communications of ACM, 1997, 40: 43~52

华中科技大学数字媒体实验室

针对特定视频类型视频分割

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

- ❖ **领域知识**
 - 新闻报道中的场景检测
 - 体育视频
 - 纪录片
 - 访谈节目
- ❖ **结合音频和视觉信息的视频分割**

华中科技大学数字媒体实验室

Q & A

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

华中科技大学数字媒体实验室

