一、图像与视频的区别

静止的图片称为**图像**（Image），由图像序列构成的数据称为**视频**（Video）

二、视频帧并非独立

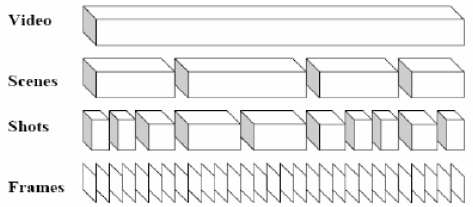
①**时间联系**：连续视频帧中特征是渐变的；②**空间联系**：连续视频帧中某些子块保持不变，某些子块却变化；

三、视频需要分割

**图像分割**一般是将图像前景和背景切分开来；

**视频分割**是将视频中的连续图像帧在特征突变地方分割开来，当然，视频分割不是将视频数据流还原成图像帧。在视频流数据中，其最小的物理数据单元是“**镜头**”，所以视频分割的目的是将连续图像帧分割成长短不一的视频镜头，而不是将视频数据流解码成一个个图像帧。

视频片段的构成：



**Scene:**Collections of shots forming a semantic(语义上的) unity.Conceptually(概念性地)，a single time and place.

**Shot:**Sequence of frames recorded in a single camera operation.

**Frame:**A single still image from a video

四、视频结构化

**帧(Frame)：**帧是视频流中的基本组成单元。每一帧均可看成一个独立的图像。①视频流数据就是由连续图像帧构成的。②在PAL视频格式中，视频帧采样率为25帧/秒，NTSC制式中，视频采样率为30帧/秒。

**镜头(Shot)：**摄像机拍下的不间断帧序列，是视频数据流进一步结构化的基础结构层。如，在拍摄“飞机起飞”这组镜头时，画面的色彩和纹理等图像特征将保持不变。①镜头是对视频流进行处理的最小物理单元，而视频帧是视频流的基本单元，它蕴涵了少许的语义内容。②由于在同一组镜头中，属于同一组镜头的图像帧之间的特征保持稳定，如果相邻图像帧之间的特征发生了明显变化，认为发生了镜头变化，需要对视频数据进行切分。

**关键帧(Key Frame):**关键帧是可以用来代表镜头内容的图像。①在切分出来镜头结构后，关键帧被用来代表各镜头的特征，进行进一步的结构化。②在一组视频镜头中，一般关键帧数目远远小于镜头所包含的图像帧数目，如对于3秒长、采样率为30帧/秒的一组视频镜头，其最简单的关键帧提取方法是把这个镜头的第一帧和最后一帧作为关键帧，所以这个镜头就使用首尾两幅图像帧来进行表示，与其所包含的90个图像帧相比，使用关键帧去表示这个视频镜头显得十分简单和有效。

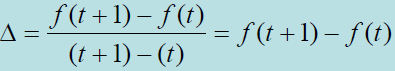
**场景(Scene)：**语义上相关和时间上相邻的若干组镜头组成了一个场景，场景是视频所蕴涵的高层抽象概念和语义的表达。如，“学校运动会”这个场景可以由“运动员入场”、“运动员比赛”和“观众呐喊”等若干镜头组成，虽然每个镜头所代表的语义不多，但是若干镜头所组合成的场景就表达了一个合符人们思维的比较丰富的语义。——由于镜头是由关键帧组成的，所以场景可以使用属于这个场景的若干镜头所对应的关键帧来表示。

**组(Group)：**组是介于物理镜头和语义场景之间的结构。例如：一段采访录像，镜头在主持人与被采访者之间频繁切换，整个采访属于一个场景，而那些关于主持人的所有镜头属于一组，关于被采访者的所有镜头属于另一组。

四、视频特征

以前面介绍的图象特征为主，反映的是**静态特性**；

如果相邻视频帧特征相减，就是特征一阶差分，体现了视频**动态变化**；

也可以按照这个思路求取视频的任意阶差分（视频的动态变化）

视频结构化的一个任务就是要把分镜头的边界自动检测出来。

五、shot detection

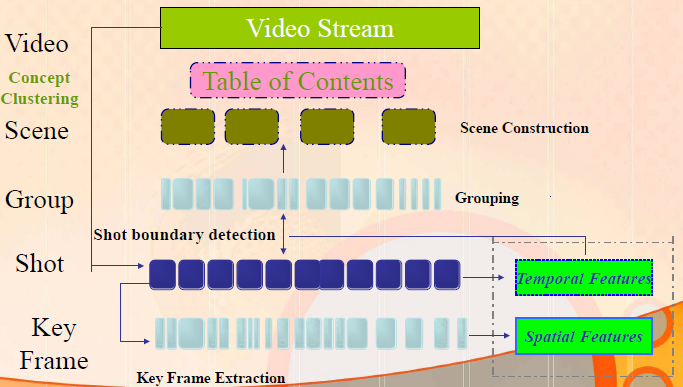
given a video V consisting of n shots, find the beginning and end of each shot. Also known as shot **boundary detection** or **transition detection.**

It is fundamental to any kind of video analysis and video application since it enables segmentation(分割) of a video into its basic components(成分): the shots.

视频镜头边界检测：视频镜头单元的自动切分与关键帧提取

视频结构生成：相似镜头聚类，形成视频目录结构

视频结构化示意图：



六、Video Shot Classification

**Hard cuts:** A cut is an instantaneous(瞬间的) transition from one scene to the next.

**Fades:** A fade is a gradual transition between a scene and a constant image (fade-out) or between a constant image and a scene (fade-in).——During a fade, images have their intensities(强度)multiplied by some value α. During a fade-in, αincreases from 0 to 1, while during a fade-out αdecreases from 1 to 0.

**Dissolves:**（溶解）A dissolve is a gradual transition from one scene to another, in which the first scene fades out and the second scene fades in.——Combination of fade-in and fade-out.

**Wipe：**（擦除）another common scene break is a wipe, in which a line moves across the screen, with the new scene appearing behind the line.



Priori：优先的

**视觉特征和转换度量是重要因素**

七、镜头边缘检测常用算法

**绝对帧间差法：**判断相邻图像帧之间特征的绝对差是否大。具体实现时，判断两个相邻帧差别的方法可以是：计算相邻两个图像帧中所有像素的色彩亮度之和，两帧的差别就定义为各自对应像素的亮度和之差。

**图象象素差法：**判断相邻图像帧中像素点发生变化的多少，达到视频镜头边缘检测的目的，这是基于图像像素差法进行镜头边缘检测的基础。首先统计两幅图像对应像素变化超过阈值的像素点个数。然后，将变化的像素点个数与第二个预定的阈值比较，如超过范围，则认为这两帧之间发生较大变化，判断其为镜头边界。但该法对镜头移动十分敏感，对噪声的容错性较差。

**图象数值差法：**不象图像像素差法比较的是相邻图像帧中所有对应像素点的特征差别，图像数值差方法将图像分成若干个子块区域，在这些区域中分别比较对应像素数值上的差别。

**颜色直方图法：**图像颜色直方图特征在镜头边缘检测中被经常使用，而基于颜色直方图特征的这种镜头边缘检测就叫做颜色直方图法。

**压缩域差法：**在这种方法中，不对图像解压，而是直接用JPEG压缩图像帧的DCT系数作为帧相似度衡量的标准。

一般来说，在对视频图像处理时候，要对图像解压，因为传统的颜色等特征是定义在像素基础上的，也就是非压缩域基础上的。

这样，基于压缩域的视频镜头检测可以省去解压步骤，直接从原始视频数据流中提取特征，从而加快检测速度。

由于使用压缩域对视频进行编码，每个压缩域系数保留了原始图像帧中或图像帧间最重要特性，所以压缩域系数可以有效分析视频数据。

**矩不变量法：**图像矩不变量具有比例、旋转和过渡不变性的特点，是用来表示图像帧的好方法，所以可以用来进行镜头边缘检测

**边界跟踪法：**在镜头的转换中，在距离原来边缘很远的位置会出现新的边缘，而原来的边缘会逐渐消失。因此，镜头转换的判断可以看作是两个图像帧中边缘的比较。

计算相邻两帧间进入或者离开图像的边所占百分比，百分比最大的是镜头的切分点。

**运动矢量法：**计算相邻视频帧之间的运动位移。

这个运动位移可以是**光流场（opticalflow）计算**，也可以**子块匹配（blockmatching）**等方法。

八、镜头边缘检测算法的实质

如何找到一种或几种良好视频图像特征，通过判断相邻图像帧之间的特征是否发生剧烈变化，来完成视频镜头边缘检测任务。

首先需要去定义特征

然后需要基于这样的特征定义其相似度函数。

九、Twin-comparison approach（双阈值）

单阈值效果不明显时

十、视频关键帧提取

视频数据流中的图像帧之间存在时间和空间冗余度——可以从视频中找一些代表性的帧(即关键帧)，用这些少量的帧来代表冗长的视频数据流，使用户看过关键帧后，就能知道了解整个视频数据流所蕴涵的内容，再通过提取这些帧的底层信息来建立索引，方便用户对视频内容查询。

十一、视频关键帧提取算法

**基于镜头边界法**

在这种方法中，将切分得到镜头中的第一幅图像和最后一幅图像作为镜头关键帧。

来自这样的观察和假设：既然在一组镜头中，相邻图像帧之间的特征变化很少，所以整个镜头中图像帧的特征变换也应该不大，因此选择镜头第一帧和最后一帧可以将镜头内容完全表达出来。

这种方法虽简单，但它不考虑当前镜头视觉内容的复杂性，并且限制了镜头关键帧的个数，使长短不同和内容不同的视频镜头都有相同个数关键帧，这样做并不合理，事实上首帧或尾帧往往并非关键帧，不能精确地代表镜头信息。

**基于特征转变法**

在基于视频图像特征提取关键帧方法中，镜头当前帧与最后一个判断为关键帧的图像比较，如有较多特征发生改变，则当前帧为新的一个关键帧。

在实际中，可以将视频镜头第一帧作为关键帧，然后比较后面视频帧图像与关键帧的图像特征是否发生了较大变化，逐渐得到后续关键帧

**基于运动分析法**

在视频摄影中，摄像机运动所造成的显著运动信息是产生图像变化的重要因素，也是提取关键帧的一个依据。

在这种方法中，将相机运动造成的图像变化分成两类：一类是相机焦距变化造成的；一类是相机角度变化造成的。

对前一种，选择首、尾两帧为关键帧；对后一种，如当前帧与上一关键帧重叠小于30%，则选其为关键帧。

**基于聚类的关键帧提取**

聚类方法在人工智能、模式识别和语音识别等领域中有着很广泛的应用，也可以使用聚类方法来提取镜头关键帧。

一般我们采取K平均聚类方法（K-Mean）

**聚类的优点**

上述聚类算法本质一个**非监督(Un-Supervised)**过程，用户只要提供聚类参数，那么聚类与关键帧提取等过程就可以自动完成，不需要人工干预，因此效率较高。

当然，在聚类过程中也可以加入些约束条件，如每个聚类集中的图像帧数目不应该很少，也不应该很多；每个聚类集的聚类质心不应该“相似”，也就是这些聚类质心的距离要大；还可以采用模糊聚类方法实现关键帧的提取。

十二、视频结构化生成

镜头切分

相似镜头成组

镜头组与镜头组之间的联系

十三、视频目录生成构造

如何构造视频目录（Table of Video Content，TOC ）

基于预定模型构造和基于通用模型构造

前者需要知道领域知识

步骤：

**镜头边缘检测→关键帧提取→时空特征提取→时间可适性成组→场景结构构造**

1. 边缘检测和关键帧提取：采取特征突变地方作为镜头边缘；每个镜头的首尾两帧作为关键帧

**时间可适性成组方法**

先计算镜头间的颜色相似度

再计算镜头间的运动相似度

最后计算总相似度

**镜头成组过程**

本质上说，是在定义两个镜头之间的相似性；

镜头中可以提取的特征很多，这里主要是使用了颜色和运动特征；

对于前面通过镜头边缘检测分割出来的镜头单元，分别计算镜头与镜头之间的总相似度，如果某两个镜头单元的总相似度值超过了事先设置的阈值，则认为这两个镜头单元应该位于同一组。

因为时间间隔相隔太远的镜头单元应该不相似的。所以在实际计算中，可以设置一个窗口（MemoryWindows），只是计算位于这个窗口内的视频单元，加快相似度比较速度。

**视频场景构造算法描述**

输入：视频镜头序列

输出：含有场景、组和镜头的视频结构

十四、视频时序结构图构造

视频解码—视频切分—关键帧提取—视频聚类分析—构造时序图—按照时序图浏览（Hyper Media）

视频例子之间的相似度计算一般采用**弹性度量算法：**Elastic matching ，本质上是比较包含元素数目不同的集合之间相似性；弹性匹配是一种采用动态时间变形（Dynamic Time Warp，DTW）技术的非线性匹配算法，已成功地运用于语音识别、签名认证和手写体识别等领域；

通过**高斯聚类**可以将相似镜头合并到一起。

**混合高斯聚类**：更好模拟同一类别中数据的差异性。

十五、视频压缩技术

Blabla来不及整理了—,—

比较重要的就是那个DCT吧

**音乐**

**基本要素：**1**音色**：即声音的感觉特性，音色是由发音物体所产生的谐波决定的，发音物体的性质、形状不同，所产生的谐波也不同，因而音色也不同 2**音高**：指各种不同高低的声音，由发音体的振动频率决定3**音强**：即音的强弱（响亮）程度，由发音时发音体振动幅度（简称振幅）的大小决定的4**音长**：是指声音的长短，它决定于发音体振动时间的久暂

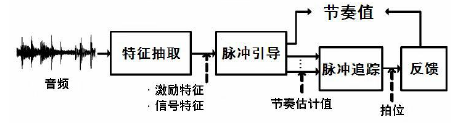
音乐计算机表示：1乐谱版面描述语言2数字乐器接口3音频信号表示

**音乐表现形式**：**旋律**，旋律是指长短、高低、强弱不同的一连串乐音有组织地进行。旋律又是曲调，它是音乐的基础和灵魂 **节奏**：节奏是指组织起来的音的长短关系，它是音乐的骨架。节拍是指时值相等的强拍和弱拍有规律地交替出现。 “有板有眼” **和声**，和声包括“和弦”及“和声进行”。和弦通常是由三个或三个以上的乐音按一定的法则纵向（同时）重叠而形成的音响组合。和弦的横向组织就是和声进行。

音乐分析：

* 基于MIDI的3D动画场景生成
  + 用户输入MIDI文件
  + 读取MIDI文件，统计音轨数与音轨相关乐器
  + 创建一个3D场景，根据音乐文件中的乐器在场景中摆放相应的物体
  + 随着音乐播放，物体随音乐节奏做出不同动作，生成动画

节奏识别框架：



节奏检测

节奏特征提取，时域分析