# 音视频中你不得不知道的知识点！

## **前言**

今天给大家分享一些音视频里面非常基础的一些概念知识，而且这些基本概念你不得不了解。下面我们快来看看吧

## **基本概念：**

### **1、编解码**

编解码器(codec)指的是 一个能够对一个信号 或者 一个数据流进行变换的 设备 或者程序。这里指的变换既包括将信号或者数据流进行编码(通常是为了传输、存储或者加密)或者提取得到一个编码流的操作，也包括 为了观察 或 处理 从这个编码流中恢复适合观察 或 操作的形式的操作。编解码器经常用在视频会议和流媒体等应用中。

### **2、容器**

很多多媒体数据流需要同时包含音频数据和视频数据，这时通常会加入一些用于音频和视频数据同步的元数据，例如字幕。这三种数据流可能会被不同的程序，进程或者硬件处理,但是当它们传输或者存储的时候，这三种数据通常是被封装在一起的。通常这种封装是通过视频文件格式来实现的，例如常见的\*.mpg，\*.avi，\*.mov， \*.mp4， \*.rm，\*.ogg or \*.tta.这些格式中有些只能使用某些编解码器，而更多可以以容器的方式使用各种编解码器。FourCC全称Four-Character Codes，是由4个字符(4 bytes)组成，是一种独立标示视频数据流格式的四字节，在wav、avi 档案之中会有一段FourCC来描述这个AVI档案, 是利用何种codec来编码的。因此wav、avi 大量存在等于“IDP3”的FourCC。视频是现在电脑中多媒体系统中的重要一环。为了适应储存视频的需要，人们设定了不同的视频文件格式来把视频和音频放在一个文件中，以防便同时回放。视频档实际上都是一个容器里面包裹着不同的轨道，使用的容器的格式关系到视频档的可扩展性。

## **参数介绍：**

### **1、采样率**

采样率(也称为采样速度或者采样频率)定义了每秒从连续信号中提取并组成离散信号的采样个数，它用赫兹(Hz)来表示。采样频率的倒数叫作采样周期或采样时间，它是采样之间的时间间隔。注意不要将采样率与比特率(bit rate，亦称“位速率")相混淆。采样定理表明采样频率必须大于被采样信号带宽的两倍，另外一种等同的说法是奈奎斯特频率必须大于被采样信号的带宽。如果信号的带宽是100Hz, 那么为了避免混叠现象采样频率必须大于200Hz。换句话说就是采样频率必须至少是信号中最大频率分量频率的两倍,否则就不能从信号采样中恢复原始信号。

对于语音采样:

8,000 Hz-电话所用采样率，对于人的说话已经足够

11,025 Hz

22,050 Hz-无线电广播所用采样率

32,000 Hz - miniDV数码视频camcorder、 DAT (LP mode)所用采样率

44,100Hz-音频CD,也常用于MPEG-1音频(VCD, SVCD, MP3)所用采样率

47 ,250 Hz - Nippon Columbia (Denon)开发的世界上第-一个商用PCM录音机所用采样率

48,000 Hz - miniDV、数字电视、DVD、DAT、电影和专业音频所用的数字声音所用采样率

50,000 Hz-二十世纪七十年代期年识别3M和Soundstream 开发的第一款商用数字录音机所用采样率

50,400Hz-三菱X-80数字录音机所用所用采样率

96,000 或者192,000 Hz - DVD-Audio、 - -些LPCM DVD音轨、Blu-ray Disc (蓝光盘)音轨、和HD-DVD ( 高清晰度DVD)音轨所用所用采样率

2.8224 MHz- SACD、索尼和飞利浦 联合开发的称为Direct Stream Digital的1位sigma-delta modulation过程所用采样率。在模拟视频中，采样率定义为帧频和场频，而不是概念上的像素时钟。图像采样频率是传感器积分周期的循环速度。由于积分周期远远小于重复所需时间，采样频率可能与采样时间的倒数不同。

50Hz- PAL视频

60/ 1.001 Hz - NTSC视频,当模拟视频转换为数字视频的时候，出现另外-种不同的采样过程，这次是使用像素频率。一些常见的像素采样率有:13.5 MHz - CCIR 601、 D1 video

### **2、分辨率：**

分辨率，泛指量测或显示系统对细节的分辨能力。此概念可以用时间、空间等领域的量测。日常用语中之分辨率多用于图像的清晰度。分辨率越高代表图像品质越好，越能表现出更多的细节。但相对的，因为纪录的信息越多，文件也就会越大。目前个人电脑里的图像,可以使用图像处理软件，调整图像的大小、编修照片等。例如photoshop, 或是photoimpact等软件。

### **3、图像分辨率：**

用以描述图像细节分辦能力，同样适用于数字图像、胶卷图像、及其他类型图像。常用线每毫米、'线 每英时'等来衡量。通常，“分辨率"被表示成每-一个方向.上的像素数量，比如：640x480等。而在某些情况下，它也可以同时表示成“每英时像素”( pixels per inch, ppi)以及图形的长度和宽度。比如72ppi,和8x6英时。

### **4、视频分辨率：**

各种电视规格分辨率比较视频的画面大小称为“分辨率”。数位视频以像素为度量单位，而类比视频以水平扫瞄线数量为度量单位。标清电视频号分辨率为720/704/640x480i60 (NTSC)或768/720x576i50 (PAL/SECAM)。新的高清电视(HDTV)分辨率可达1920x1080p60,即每条水平扫瞄线有1920个像素，每个画面有1080条扫瞄线，以每秒钟60张画面的速度播放。

### **5、画面更新率fps**

Framerate中文常译为“画面更新率"或“帧率”，是指视频格式每秒钟播放的静态画面数量。典型的画面更新率由早期的每秒6或8张(frame persecond,简称fps)，至现今的每秒120张不等。PAL (欧洲，亚洲，澳洲等地的电视广播格式)与SECAM (法国，俄国，部分非洲等地的电视广播格式)规定其更新率为25fps，而NTSC (美国，加拿大，日本等地的电视广播格式)则规定其更新率为29.97 fps。电影胶卷则是以稍慢的24fps在拍摄，这使得各国电视广播在播映电影时需要-些复杂的转换手续(参考Telecine转换)。要达成最基本的视觉暂留效果大约需要10fps的速度。

### **6、压缩方法：**

* 有损压缩和无损压缩

在视频压缩中有损(Lossy )和无损(Lossless) 的概念与静态图像中基本类似。

无损压缩也即压缩前和解压缩后的数据完全-致。多数的无损压缩都采用RLE行程编码算法。

有损压缩意味着解压缩后的数据与压缩前的数据不一致。在压缩的过程中要丢失一些人眼和人耳所不敏感的图像或音频信息，而且丢失的信息不可恢复。几乎所有高压缩的算法都采用有损压缩，这样才能达到低数据率的目标。丢失的数据率与压缩比有关，压缩比越小，丟失的数据越多，解压缩后的效果一般越差。此外，某些有损压缩算法采用多次重复压缩的方式，这样还会引起额外的数据丢失。

无损格式，例如WAV, PCM，TTA，FLAC, AU, APE，TAK, WavPack(WV)

有损格式，例如MP3, Windows Media Audio (WMA)，Ogg Vorbis (OGG)，AAC

帧内压缩 和 帧间压缩：

帧内(Intraframe) 压缩也称为空间压缩(Spatial compression)。当压缩一帧图像时，仅考虑本帧的数据而不考虑相邻帧之间的冗余信息，这实际上与静态图像压缩类似。帧内一般采用有损压缩算法，由于帧内压缩时各个帧之间没有相互关系，所以压缩后的视频数据仍可以以帧为单位进行编辑。帧内压缩一般达不到很高的压缩。

采用帧间(Interframe )压缩是基于许多视频或动画的连续前后两帧具有很大的相关性，或者说前后两帧信息变化很小的特点。也即连续的视频其相邻帧之间具有冗余信息，根据这一特性，压缩相邻帧之间的冗余量就可以进一步提高压缩量，减小压缩比。帧间压缩也称为时间压缩(Temporalcompression) ，它通过比较时间轴上不同帧之间的数据进行压缩。帧间压缩一般是无损的。帧差值( Frame differencing)算法是一种典型的时间压缩法，它通过比较本帧与相邻帧之间的差异，仅记录本帧与其相邻帧的差值，这样可以大大减少数据量。

* 对称编码和不对称编码

对称性(symmetric) 是压缩编码的一个关键特征。对称意味着压缩和解压缩占用相同的计算处理能力和时间，对称算法适合于实时压缩和传送视频，如视频会议应用就以采用对称的压缩编码算法为好。而在电子出版和其它多媒体应用中，一般是把视频预先压缩处理好，然后再播放，因此可以采用不对称(asymmetric) 编码。不对称或非对称意味着压缩时需要花费大量的处理能力和时间，而解压缩时则能较好地实时回放，也即以不同的速度进行压缩和解压缩。一般地说，压缩一段视频的时间比回放(解压缩)该视频的时间要多得多。例如，压缩一段三分钟的视频片断可能需要10多分钟的时间，而该片断实时回放时间只有三分钟。

## **什么是I帧、B帧、P帧：**

视频压缩中，每帧代表一幅静止的图像。而在实际压缩时，会采取各种算法减少数据的容量，其中IPB就是最常见的。

简单地说：

I帧是关键帧，属于帧内压缩。就是和AVI的压缩是一样的。

P是向前搜索的意思。

B是双向搜索。

他们都是基于I帧来压缩数据。

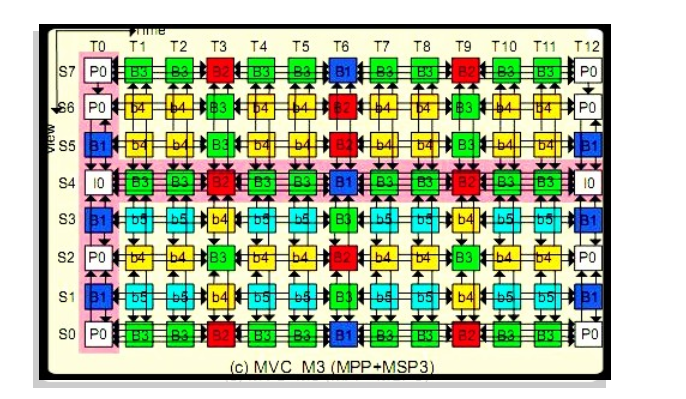
I帧表示关键帧，你可以理解为这一帧画面的完整保留；解码时只需要本帧数据就可以完成（因为包含完整画面）

P帧表示的是这一帧跟之前的一个关键帧（或P帧）的差别，解码时需要用之前缓存的画面叠加上本帧定义的差别，生成最终画面。（也就是差别帧，P帧没有完整画面数据，只有与前一帧的画面差别的数据）

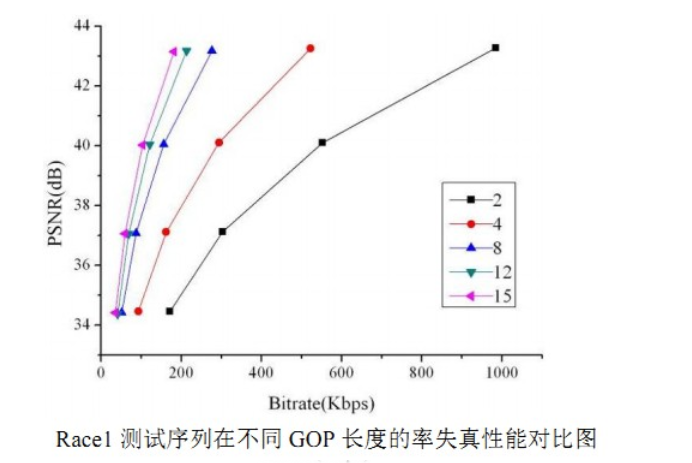
B帧是双向差别帧，也就是B帧记录的是本帧与前后帧的差别（具体比较复杂，有4种情况），换言之，要解码B帧，不仅要取得之前的缓存画面，还要解码之后的画面，通过前后画面的与本帧数据的叠加取得最终的画面。B帧压缩率高，但是解码时CPU会比较累~。

从上面的解释看，我们知道I和P的解码算法比较简单，资源占用也比较少，I只要自己完成就行了，P呢，也只需要解码器把前一个画面缓存一下，遇到P时就使用之前缓存的画面就好了，如果视频流只有I和P，解码器可以不管后面的数据，边读边解码，线性前进，大家很舒服。但网络上的电影很多都采用了B帧，因为B帧记录的是前后帧的差别，比P帧能节约更多的空间，但这样一来，文件小了，解码器就麻烦了，因为在解码时，不仅要用之前缓存的画面，还要知道下一个I或者P的画面（也就是说要预读预解码），而且，B帧不能简单地丢掉，因为B帧其实也包含了画面信息，如果简单丢掉，并用之前的画面简单重复，就会造成画面卡（其实就是丢帧了），并且由于网络上的电影为了节约空间，往往使用相当多的B帧，B帧用的多，对不支持B帧的播放器就造成更大的困扰，画面也就越卡。  一般平均来说，I的压缩率是7（跟JPG差不多），P是20，B可以达到50，可见使用B帧能节省大量空间，节省出来的空间可以用来保存多一些I帧，这样在相同码率下，可以提供更好的画质。

下面举例说明：



在如上图中，GOP (Group of Pictures)长度为13，S0~S7 表示 8个视点，T0~T12 为 GOP的 13个时刻。每个 GOP包含帧数为视点数 GOP 长度的乘积。在该图中一个 GOP 中，包含94 个 B帧。B 帧占一个 GOP 总帧数的 90.38%。GOP 越长，B 帧所占比例更高，编码的率失真性能越高。下图测试序列 Race1 在不同 GOP 下的率失真性能对比。



## **总结：**

以上就是今天的分享，如果对音视频感兴趣的朋友，可以扫码加我个人微信；