音视频知识总结

音视频技术主要包括以下几点：封装技术、视频压缩编码技术、音频压缩编码技术，如果考虑到网络传输的话，还包括流媒体协议技术。

视频播放器原理：

如果要播放一个互联网上的视频文件，需要经过一下几个步骤：解协议、解封装、解码视音频、视音频同步

解协议：将流媒体协议的数据解析为标准的相应的封装格式数据。音视频在网络上传播时，常常会采用各种流媒体协议，例如HTTP、RTMP等，解协议过程中回去除掉信令数据而只保留音视频数据。例如RTMP协议传输的数据，经过解析后输出FLV格式的数据。

解封装：就是将输入的封装格式的数据，分离成为音频流压缩编码数据和视频流压缩编码数据。封装格式种类很多，如MP4、MKV等，他的作用就是将已经压缩编码的视频数据和音频数据按照一定的格式放到一起。例如FLV格式的数据，经过解封装操作后，输出H.264编码的视频码流和AAC编码的音频码流。

解码：就是将音频、视频压缩码流数据，解码成为非压缩的音视频原始数据。音频的压缩编码标准包含AAC、MP3等等，视频的压缩编码标准则包含H.264，MPEG2等。解码是整个系统中最重要也是最复杂的一个环节。通过解码，压缩编码的视频数据输出成为非压缩的颜色数据，如YUV420P，RGB等等。压缩编码的音频数据输出成为非压缩的音频抽样数据，例如PCM数据。

音视频同步：根据解封装模块处理过程中获取到的参数信息，同步解码出来的视频和音频数据，并将视频音频数据送至系统的显卡和声卡播放出来。

1.封装格式

不同视频文件的后缀格式(avi,rmvb,mp4,flv,mkv)不同，这些格式代表的是封装格式，就是把视频数据和音频数据打包成一个文件的规范。有些封装格式支持的视音频编码标准十分广泛，应该算比较优秀的封装格式，如MKV；而有些封装格式支持的音视频编码标准很少，应该属于落后的封装格式，如RMVB。

2.主要流媒体协议

因为互联网环境的不稳定性，互联网视频服务通常采用TCP作为其流媒体的传输层协议，因而像RTMP、HTTP这类的协议广泛用于互联网音视频服务之中。这类协议不会发生丢包，因而保证了视频的质量。

3.视频编码

视频编码的主要作用是将视频像素数据（RGB、YUV等）压缩成为视频码流，从而降低视频的数据量。如果视频不经过压缩编码的话，体积通常是非常大的。主流编码标准有H.264、MPEG4等

4.音频编码

主要作用是将音频采样数据（PCM等）压缩成为音频码流，从而降低音频的数据量。常见的编码标准有AAC、WMA和MP3等

5.网络音视频平台对比

直播服务普遍采用了RTMP作为流媒体协议，FLV作为封装格式，H.264作为视频编码格式，AAC作为音频编码格式。采用RTMP作为直播协议的好处在于齐备Flash播放器支持。

点播服务普遍采用了HTTP作为流媒体协议，H.264作为视频编码格式，AAC作为音频编码格式。采用HTTP有以下两点优势：一方面HTTP基于TCP协议的应用层协议，媒体传输过程中不会出现丢包等现象；另一方面是HTTP被绝大部分的web服务器支持，不需要投资购买额外的流媒体服务器。

视频编码基本原理

以记录数字视频的YUV分量格式为例，YUV分别代表亮度和两个色差信号。其亮度信号采样频率为13.5MHz，色度信号的频带通常为亮度信号的一半或更少，为6.75MHz或3.375MHz，Y信号采用13.5MHz，色度信号U和V采用6.75MHz采样，采样信号以8bit量化，可以计算出数字视频的码率为

13.5\*8+6.75\*8+6.75\*8=216Mbit/s

如此大的数据很难直接进行存储或传输，所以必须采用压缩技术以减少码率。一般压缩编码方法都是混合编码，即将变换编码、运动估计和运动补偿，以及熵编码三种方式相结合。

1. 变换编码

将空间域描述的图像信号变换到频率域，然后对变换后的系数进行编码处理。一般来说，图像在空间上具有较强的相关性，变换到频率域可以实现去相关和能量集中。

1. 熵编码

因编码后的平均码长接近信源熵值而得名。

1. 运动估计和运动补偿

消除图像序列时间方向相关性的有效手段。变换编码和熵编码是在一帧图像的基础上进行的，通过这些方法可以消除图像内部各像素间在空间上的相关性。实际上图像信号除了空间上的相关性之外，还有时间上的相关性。编码器将输入的每一帧图像根据其参考图像的不同分成3种不同的类型：I(Intra)帧、B(Bidirection prediction)帧、P(Prediction)帧。I帧只使用本帧内的数据进行编码，压缩比相对不高。P帧在编码过程中使用一个前面的I帧或P帧作为参考进行运动补偿，实际上是对当前图像和参考图像的差值进行编码。B帧的编码方式和P帧相似，唯一不同的是在编码过程中它要使用一个前边的I/P或后边的I/P帧进行预测，相比之下，B帧比P帧拥有更高的压缩比。

1. 混合编码

把上面三种结合起来进行编码

音频编码基本原理

1. 音频信号的冗余信息

冗余信息包含人耳听觉范围外的音频信号以及被掩蔽掉的音频信号等。例如，人耳察觉到的频率范围为20Hz~20KHz，除此之外都可视为冗余信号；当一个强音信号和一个弱音信号同时存在时，弱音将被掩蔽。

1. 压缩编码方法

对每一个音频声道中的音频采样信号，首先都要将他们映射到频域中，这种时域到频域的映射可通过子带滤波器实现。

H.264视频码流数据解析

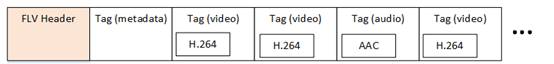
H.264原始码流是由一个一个的NALU组成的，其中每个NALU之间通过startcode(起始码)进行分隔，起始码分成两种：0x000001或者0x00000001。如果NALU对应的Slice为一帧的开始就用0x00000001，否则就用0x000001。H.264码流解析的步骤就是首先从码流中搜索0x000001和0x00000001，分理处NALU；然后再分析NALU的各个字段。

AAC音频码流解析

AAC原始码流是由一个一个的ADTS frame组成的，其中每个ADTS frame之间通过syncword(同步字)进行分隔。同步字为0xFFF（二进制“111111111111”）。AAC码流解析的步骤就是首先从码流中搜索0x0FFF，分流出ADTS frame；然后再分析ADTS frame的首部各个字段。

FLV封装格式解析

FLV封装格式是由一个FLV Header文件头和一个一个的Tag组成的。Tag中包含了音频数据以及视频数据，如下如所示



其中，每个Tag前面还包含了Previous Tag Size字段，表示前面一个Tag的大小。Tag类型可以使视频、音频和Script，每个Tag只能包含以上三种类型的数据中的一种。