

一直想把音视频编解码技术做一个简单的总结，可是苦于时间不充裕，一直没能完成。今天有着很大的空闲，终于可以总结一个有关音视频技术的入门教程，可以方便更多的人学习从零开始学习音视频技术。需要注意的是，本文所说的音视频技术，指的是理论层面的音视频技术，并不涉及到编程相关的东西。

0. 生活中的音视频技术

平时我们打开电脑中自己存电影的目录的话，一般都会如下图所示，一大堆五花八门的电影。（其实专业的影视爱好者一概会把影视文件分门别类的，但我比较懒，一股脑把电影放在了一起）



因为下载的来源不同，这些电影文件有不同的格式，用不同的后缀表示：avi, rmvb, mp4, flv, mkv等等（当然也使用不同的图标）。在这里需要注意的是，这些格式代表的是**封装格式**。何为封装格式？就是把视频数据和音频数据打包成一个文件的规范。仅仅靠看文件的后缀，很难能看出具体使用了什么**视音频编码标准**。总的来说，不同的封装格式之间差距不大，各有优劣。

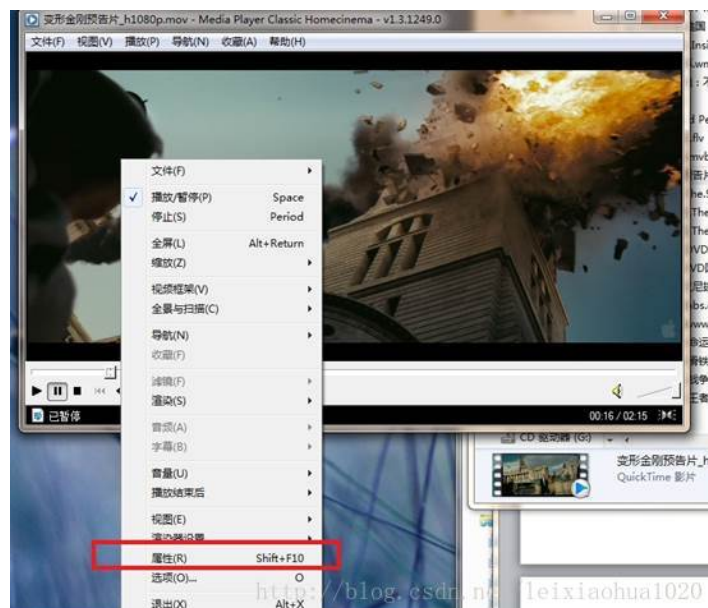
注：有些封装格式支持的视音频编码标准十分广泛，应该算比较优秀的封装格式，比如MKV；而有些封装格式支持的视音频编码标准很少，应该属于落后的封装格式，比如RMVB。

下面演示一下如何使用视频播放器查看一个视频文件采用的视音频技术：

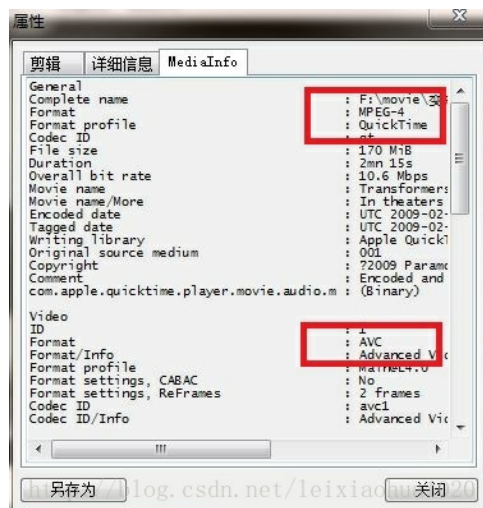
注：在这里使用Media Player Classic作为演示，实际上暴风影音，KMplayer这些操作都差不多。

1. 打开一个视频文件

2. 在视频画面上右键单击，选择属性。



3. 在弹出的“属性”对话框的“MediaInfo”选项卡中，就可以查看该视频文件采用的视音频技术了。如下图所示，“变形金刚预告片_h1080p.mov”采用MPEG4的QuickTime封装格式（MOV），采用了H.264（AVC）的压缩编码标准。



在这里用到了一个查看媒体信息的工具：MediaInfo。MediaInfo是一个专门查看视音频格式的工具，软件的详细使用可参考：

[MediaInfo 使用简介（新版本支持 HEVC）](#)

源代码分析可参考：

[MediaInfo 源代码分析 1：整体结构 \[系列文章 \]](#)

生活中的视音频技术到此为止，下面介绍一下具体的视音频技术。

1. 视频播放器原理

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

视频播放器的源代码详细解析（Media Player Classic - HC, Mplayer, FFplay, XBMC）可以参考系列文章：

[Media Player Classic：Media Player Classic - HC 源代码分析 1：整体结构 \[系列文章 \]](#)

[Mplayer：MPlayer 源代码分析](#)

[FFplay：FFplay 源代码分析：整体流程图](#)

[XBMC：XBMC 源代码分析 1：整体结构以及编译方法 \[系列文章 \]](#)

在这里不细说了，仅简要说明一下视频播放器的原理。

视频播放器播放一个互联网上的视频文件，需要经过以下几个步骤：解协议，解封装，解码视音频，视音频同步。如果播放本地文件则不需要解协议，为以下几个步骤：解封装，解码视音频，视音频同步。他们的过程如图所示。



解协议的作用，就是将流媒体协议的数据，解析为标准的相应的封装格式数据。视音频在网络上传播的时候，常常采用各种流媒体协议，例如HTTP, RTMP, 或是MMS等等。这些协议在传输视音频数据的同时，也会传输一些信令数据。这些信令数据包括对播放的控制（播放，暂停，停止），或者对网络状态的描述等。解协议的过程中会去除掉信令数据而只保留视音频数据。例如，采用RTMP协议传输的数据，经过解协议操作后，输出FLV格式的数据。

解封装的作用，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，例如MP4, MKV, RM

VB, TS, FLV, AVI等等，它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 的作用，就是将视频/音频压缩编码数据，解码成为非压缩的视频/音频原始数据。音频的压缩编码标准包含AAC，MP3，AC-3等等，视频的压缩编码标准则包含H.264，MPEG2，VC-1等等。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，例如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。

视音频同步 的作用，就是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。

接下来的几节我们将会列出主要的流媒体协议，封装格式，以及视音频编码标准。更详细的比较可以参考：

[视频参数（流媒体系统，封装格式，视频编码，音频编码，播放器）对比](#)

有关本文中涉及到的协议数据、封装格式数据、视频编码数据、音频编码数据、视频像素数据、音频采样数据的分析可以参考下面系列文章：

[视音频数据处理入门：RGB、YUV像素数据处理](#)

[视音频数据处理入门：PCM音频采样数据处理](#)

[视音频数据处理入门：H.264视频码流解析](#)

[视音频数据处理入门：AAC音频码流解析](#)

[视音频数据处理入门：FLV封装格式解析](#)

[视音频数据处理入门：UDP-RTP协议解析](#)

2. 流媒体协议

流媒体协议是服务器与客户端之间通信遵循的规定。当前网络上主要的流媒体协议如表所示。

主要流媒体协议一览

名称	推出机构	传输层协议	客户端	目前使用领域
RTSP+RTP	IETF	TCP+UDP	VLC, WMP	IPTV
RTMP	Adobe Inc.	TCP	Flash	互联网直播
RTMFP	Adobe Inc.	UDP	Flash	互联网直播
MMS	Microsoft Inc.	TCP/UDP	WMP	互联网直播+点播
HTTP	WWW+IETF	TCP	Flash	互联网点播

RTSP+RTP经常用于IPTV领域。因为其采用UDP传输视音频，支持组播，效率较高。但其缺点是网络不好的情况下可能会丢包，影响视频观看质量。因而围绕IPTV的视频质量的研究还是挺多的。

RTSP规范可参考：[RTSP 协议学习笔记](#)

RTSP+RTP系统中衡量服务质量可参考：[网络视频传输的服务质量（QoS）](#)

上海IPTV码流分析结果可参考：[IPTV 视频码流分析](#)

因为互联网网络环境的不稳定性，RTSP+RTP较少用于互联网视音频传输。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，MMS，HTTP这类的协议广泛用于互联网视音频服务之中。这类协议不会发生丢包，因而保证了视频的质量，但是传输的效率会相对低一些。

此外RTMFP是一种比较新的流媒体协议，特点是支持P2P。

RTMP我做的研究相对多一些：比如 [RTMP 规范简单分析](#)，或者 [RTMP 流媒体播放过程](#)

相关工具的源代码分析：[RTMPdump 源代码分析 1：main\(\) 函数 \[系列文章 \]](#)

RTMP协议学习：[RTMP 流媒体技术零基础学习方法](#)

3. 封装格式

封装格式的主要作用是把视频码流和音频码流按照一定的格式存储在一个文件中。现如今流行的封装格式如下表所示：

主要封装格式一览

名称	推出机构	流媒体	支持的视频编码	支持的音频编码	目前使用领域
AVI	Microsoft Inc.	不支持	几乎所有格式	几乎所有格式	BT下载影视
MP4	MPEG	支持	MPEG-2, MPEG-4, H.264, H.263等	AAC, MPEG-1 Layers I, II, III, AC-3等	互联网视频网站
TS	MPEG	支持	MPEG-1, MPEG-2, MPEG-4, H.264	MPEG-1 Layers I, II, III, AAC,	IPTV, 数字电视
FLV	Adobe Inc.	支持	Sorenson, VP6, H.264	MP3, ADPCM, Linear PCM, AAC等	互联网视频网站
MKV	CoreCodec Inc.	支持	几乎所有格式	几乎所有格式	互联网视频网站
RMVB	Real Networks Inc.	支持	RealVideo 8, 9, 10	AAC, Cook Codec, RealAudio Lossless	BT下载影视

由表可见，除了AVI之外，其他封装格式都支持流媒体，即可以“边下边播”。有些格式更“万能”一些，支持的视音频编码标准多一些，比如MKV。而有些格式则支持的相对比较少，比如说RMVB。

这些封装格式都有相关的文档，在这里就不一一例举了。

我自己也做过辅助学习的小项目：

[TS 封装格式分析器](#)

[FLV 封装格式分析器](#)

4. 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，一部电影可能就要上百G的空间。视频编码是视音频技术中最重要的技术之一。视频码流的数据量占了视音频总数据量的绝大部分。高效率的视频编码在同等的码率下，可以获得更高的视频质量。

视频编码的简单原理可以参考：[视频压缩编码和音频压缩编码的基本原理](#)

注：视频编码技术在整个视音频技术中应该是最复杂的技术。如果没有基础的话，可以先买一些书看一下原理，比如说《现代电视原理》《数字电视广播原理与应用》（本科的课本）中的部分章节。

主要视频编码一览

名称	推出机构	推出时间	目前使用领域
HEVC(H.265)	MPEG/ITU-T	2013	研发中
H.264	MPEG/ITU-T	2003	各个领域
MPEG4	MPEG	2001	不温不火
MPEG2	MPEG	1994	数字电视
VP9	Google	2013	研发中
VP8	Google	2008	不普及
VC-1	Microsoft Inc.	2006	微软平台

由表可见，有两种视频编码方案是最新推出的：VP9和HEVC。目前这两种方案都处于研发阶段，还没有到达实用的程度。当前使用最多的视频编码方案就是H.264。

4.1 主流编码标准

H.264仅仅是一个编码标准，而不是一个具体的编码器，H.264只是给编码器的实现提供参照用的。

基于H.264标准的编码器还是很多的，究竟孰优孰劣？可参考：[MSU 出品的 H.264 编码器比较（2011.5）](#)

在学习视频编码的时候，可能会用到各种编码器（实际上就是一个exe文件），他们常用的编码命令可以参考：[各种视频编码器的命令行格式](#)

学习H.264最标准的源代码，就是其官方标准JM了。但是要注意，JM速度非常的慢，是无法用于实际的：[H.264 参考软件 JM12.2RC 代码详细流程](#)

实际中使用最多的就是x264了，性能强悍(超过了很多商业编码器)，而且开源。其基本教程网上极多，不再赘述。编码时候可参考：[x264 编码指南——码率控制](#)。编码后统计值的含义：[X264 输出的统计值的含义（X264 Stats Output）](#)

Google推出的VP8属于和H.264同一时代的标准。总体而言，VP8比H.264要稍微差一点。有一篇写的很好的VP8的介绍文章：[深入了解 VP8](#)。除了在技术领域，VP8和H.264在专利等方面也是打的不可开交，可参考文章：[WebM\(VP8\) vs H.264](#)

此外，我国还推出了自己的国产标准AVS，性能也不错，但目前比H.264还是要稍微逊色一点。不过感觉我国在视频编解码领域还算比较先进的，

可参考：[视频编码国家标准 AVS 与 H.264 的比较（节选）](#)

近期又推出了AVS新一代的版本AVS+，具体的性能测试还没看过。不过据说AVS+得到了国家政策上非常强力的支持。

4.2 下一代编码标准

下一代的编解码标准就要数HEVC和VP9了。VP9是Google继VP8之后推出的新一代标准。VP9和HEVC相比，要稍微逊色一些。它们的对比可参考：[（1）HEVC与VP9编码效率对比](#) [（2）HEVC，VP9，x264性能对比](#)

HEVC在未来拥有很大优势，可参考：[HEVC 将会取代 H.264 的原因](#)

学习HEVC最标准的源代码，就是其官方标准HM了。其速度比H.264的官方标准代码又慢了一大截，使用可参考：[HEVC 学习 —— HM 的使用](#)

未来实际使用的HEVC开源编码器很有可能是x265，目前该项目还处于发展阶段，可参考：[x265\(HEVC 编码器，基于 x264\) 介绍](#)。x265的使用可以参考：[HEVC （H.265）标准的编码器（x265，DivX265）试用](#)

主流以及下一代编码标准之间的比较可以参考文章：[视频编码方案之间的比较（HEVC，H.264，MPEG2等）](#)

此外，在码率一定的情况下，几种编码标准的比较可参考：[限制码率的视频编码标准比较（包括MPEG-2，H.263，MPEG-4，以及H.264）](#)

结果大致是这样的：

HEVC > VP9 > H.264> VP8 > MPEG4 > H.263 > MPEG2。

截了一些图，可以比较直观的了解各种编码标准：

HEVC码流简析：[HEVC 码流简单分析](#)

H.264码流简析：[H.264 简单码流分析](#)

MPEG2码流简析：[MPEG2 简单码流分析](#)

以上简析使用的工具：[视频码流分析工具](#)

我自己做的小工具：[H.264码流分析器](#)

5. 音频编码

音频编码的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。音频编码也是互联网视音频技术中一个重要的技术。但是一般情况下音频的数据量要远小于视频的数据量，因而即使使用稍微落后的音频编码标准，而导致音频数据量有所增加，也不会对视音频的总数据量产生太大的影响。高效率的音频编码在同等的码率下，可以获得更高的音质。

音频编码的简单原理可以参考：[视频压缩编码和音频压缩编码的基本原理](#)

主要音频编码一览

名称	推出机构	推出时间	目前使用领域
AAC	MPEG	1997	各个领域（新）
AC-3	Dolby Inc.	1992	电影
MP3	MPEG	1993	各个领域（旧）
WMA	Microsoft Inc.	1999	微软平台

由表可见，近年来并未推出全新的音频编码方案，可见音频编码技术已经基本可以满足人们的需要。音频编码技术近期绝大部分的改动都是在MP3的继任者——AAC的基础上完成的。

这些编码标准之间的比较可以参考文章：[音频编码方案之间音质比较（AAC，MP3，WMA等）](#)

结果大致是这样的：

AAC+ > MP3PRO > AAC> RealAudio > WMA > MP3

AAC格式的介绍：[AAC 格式简介](#)

AAC几种不同版本之间的对比：[AAC 规格（LC，HE，HEv2）及性能对比](#)

AAC专利方面的介绍：[AAC 专利介绍](#)

此外杜比数字的编码标准也比较流行，但是貌似比最新的AAC稍为逊色：[AC-3 技术综述](#)

6. 现有网络视音频平台对比

现有的网络视音频服务主要包括两种方式：点播和直播。点播意即根据用户的需要播放相应的视频节目，这是互联网视音频服务最主要的方式。绝大部分视频网站都提供了点播服务。直播意即互联网视音频平台直接将视频内容实时发送给用户，目前还处于发展阶段。直播在网络电视台，社交视频网站较为常见。

6.1 直播平台参数对比

主流互联网视音频平台直播服务的参数对比如表所示：

现有网络视音频平台参数对比

名称	协议	封装	视频编码	音频编码	播放器
CNTV	私有				
华数TV	RTMP	FLV	H.264	AAC	Flash
六间房	RTMP	FLV	H.264	AAC	Flash
中国教育电视台	RTMP	FLV	H.264	AAC	Flash
北广传媒移动电视	RTMP	FLV	H.264	AAC	Flash
上海IPTV	RTSP+RTP	TS	H.264	MP2	机顶盒

可以看出，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其被Flash播放器支持。而Flash播放器如今已经安装在全球99%的电脑上，并且与浏览器结合的很好。因此这种流媒体直播平台可以实现“无插件直播”，极大的简化了客户端的操作。封装格式，视频编码，音频编码方面，无一例外的使用了FLV + H.264 + AAC的组合。FLV是RTMP使用的封装格式，H.264是当今实际应用中编码效率最高的视频编码标准，AAC则是当今实际应用中编码效率最高的音频编码标准。视频播放器方面，都使用了Flash播放器。

6.2 点播平台参数对比

主流网络视音频平台点播服务的参数对比如表所示：

现有互联网视音频平台参数对比

名称	协议	封装	视频编码	音频编码	播放器
CNTV	HTTP	MP4	H.264	AAC	Flash
CNTV（部分）	RTMP	FLV	H.264	AAC	Flash
华数TV	HTTP	MP4	H.264	AAC	Flash
优酷网	HTTP	FLV	H.264	AAC	Flash
土豆网	HTTP	F4V	H.264	AAC	Flash
56网	HTTP	FLV	H.264	AAC	Flash
音悦台	HTTP	MP4	H.264	AAC	Flash
乐视网	HTTP	FLV	H.264	AAC	Flash
新浪视频	HTTP	FLV	H.264	AAC	Flash

可以看出，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作为点播协议有以下两点优势：一方面，HTTP是基于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，从而保证了视频的质量；另一方面，HTTP被绝大部分的Web服务器支持，因而流媒体服务机构不必投资购买额外的流媒体服务器，从而节约了开支。点播服务采用的封装格式有多种：MP4，FLV，F4V等，它们之间的区别不是很大。视频编码标准和音频编码标准是H.264和AAC。这两种标准分别是当今实际应用中编码效率最高的视频标准和音频标准。视频播放器方面，无一例外的都使用了Flash播放器。

此PDF由spygg生成,请尊重原作者版权!!!

我的邮箱:liushide@163.com