视频的春秋战国

我参加新语丝科普征文已经好多届了，最初的原因是看到第二期的榜首文章：Albert\_JIAO的《冷战，彩色电视的三国风云》感到很亲切，因为这也是本人的专业领域。后来我也陆续写了一些这方面的文章，投了几次稿。这些年视频的发展已经远远超过了电视领域，所以狗尾续貂，聊一聊电视之后的视频发展，其中的波澜壮阔，波谲云诡远远超过了区区三种电视制式的割据年代，所以就命名为视频的春秋战国吧。

人接收的信息80%以上是通过“看”来得到的。长期以来，得到视频的途径就是电影，然后是电视，彩色电视。NTSC,PAL,SECAM三种彩色制式在全球各占领了一些地盘，大家基本相安无事。

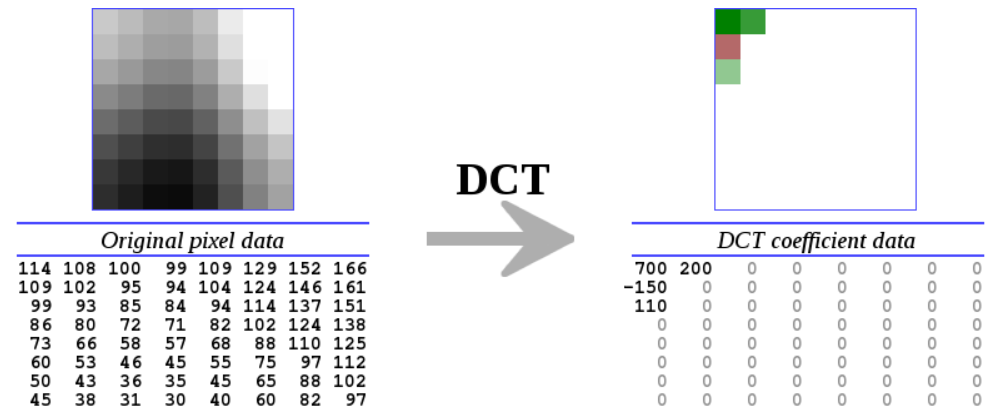
视频信息有一个特点，就是信息量特别的大。在传统的模拟电视时代，工程师们为了在有限的频带内插入尽可能多的频道，绞尽脑汁，用了很多的办法。比如用隔行扫描不影响静态图像分辨率的原理，每帧画面只传输一半的行数，这样所需的信号带宽就只有一半。另外利用人眼睛对彩色不如黑白敏感的特点，把彩色信息带宽降低，叠加在黑白信号之上，这样勉强可以实现彩色标准清晰度的画面。但拿今天的标准回头去看传统的模拟电视，简直惨不忍睹，行闪烁，爬点，混色，噪音雪花，问题非常多。

最主要的问题还是分辨率低，只有400线左右，也就是显像管上横向只能分辨出400条竖线。随着显像管尺寸越来越大，画面清晰度的提高越来越重要。各个国家纷纷憋足劲研究下一代的高清电视标准。其中80年代的电视出口大国日本最起劲，自1989年开始播放模拟高清卫星电视。如果按高清横向纵向比标清都清晰一倍来说，整个画面就是4倍的信息量。要知道在同样的带宽内把高清信号塞进去可是了不起的技术奇迹。

就在日本人踌躇满志的时候，搅局者来了。美国开始研制的数字电视技术实现了弯道超车，视频从模拟时代跨入了数字时代，揭开了视频革命的春秋战国序幕。

# 数字压缩

数字图像压缩主要是基于DCT（离散余弦变换），这个矩阵变换可以把二维图像的主要信息集中到少数项上，变成一个稀疏矩阵。对少数项的编码就能传递整个图像，大大减少了数据量。



图表 1 8x8图像的DCT ( <https://people.xiph.org/~xiphmont/demo/daala/demo1.shtml>)

图里的8x8,64个点的数据，经过DCT后，变成了右边的4个数字了。只要把这4个数发出去，接收方就能还原出原来的64个点。

大家常用的JPEG图像格式就运用了这种压缩方式。搞摄影的都知道，存原图(raw)格式要比JPEG格式大得多，一般JPEG图像可以压缩到原来的20倍左右而没有明显的质量损失。

前面说过，日本模拟高清想尽办法，比标准清晰度电视提高了4倍的信息。一用数字技术，轻轻松松就有了20倍的压缩率，革命性的变化啊。

20倍的压缩率只是刚刚起步，我们发现还有一个很大的地方可以改进，那就是丢弃冗余信息。比如一张图上有两个完全一样的物体，我们把第一个物体的图块发过去后，对第二个同样的图块要不要再发一遍呢？当然没必要，只要把第二个物体的位置发过去就可以了。

这有点像一个游戏：连连看。找同样的图块，消除冗余的信息。



图表 2 连连看，两串葡萄只要发第一串葡萄的图块就行了。

单张图压缩到连续视频压缩，能找到的相同图块就更多了。在连续视频中，一帧图像中的物体很大可能在下一帧中仍然存在。利用这个特性，我们可以大幅度地消除冗余。

A screen shot of a cat

Description automatically generated

图表 3 相邻两帧图像中的冗余部分（<https://www.researchgate.net/figure/Illustration-of-the-redundant-information-appears-in-a-video-A-The-1-st-f-rame-B_fig17_232735479>）

上图中拍摄了两张猫的视频。A和B图像中不变的背景占了绝大部分画面，这样对B图，我们只需要传递改变的猫头部分就可以了(C图)，这样所需要的信息量大大减少。

利用这个特性，对视频我们能够压缩到原来大小的200分之一！这就是大名鼎鼎的MPEG1标准。

ISO，全称International Standards Organization，即国际标准化组织。该组织下属的MPEG(Motion Picture Experts Group)，即移动图像专家组，1991年推出了MPEG1标准，这个标准被VCD采用。随后1994年推出的MPEG2基本相同但定义了对更高分辨率的支持，广泛应用于DVD,数字电视等等领域，也为大批卖碟游击队带来了商机。

碟片这种媒体是除了电视，电影之外的第三种视频媒介，迅速取代了老的模拟录像带。碟片出租业务(比如：Blockbuster)也迅速取代了录像带出租业务。

# 互联网时代

蓬勃发展的互联网开始拥有了越来越高的数据速率，视频开始在互联网上出现。主要有基于文件的视频文件和基于实时流的流媒体技术。

延续MPEG1，MPEG2的大好形势，MPEG专家组开发了MPEG4标准。

中间为何没有MPEG3呢？其实最早是有MPEG3组的，他们希望开发一套1080p高清电视的标准。后来发现MPEG2够用，于是就不用另搞一套了，这个项目也就和MPEG2组合并了。

大家不要把另一个有名的音频压缩标准MP3和MPEG3混起来。MP3实际上是MPEG1和MPEG2标准中的第三部分的意思，因为这两个标准的第一部分讲总纲，第二部分讲视频，第三部分讲音频，所以大家就简称MP3。

MPEG4比起MPEG1,2要复杂得多，主要是计算机界的程序猿们一介入，引入了封装，面向对象等等高大上的概念。带来的好处就是MPEG4可用的范围非常广泛，灵活性，可扩展性也非常强。MPEG4标准也就催生出众多的符合MPEG4标准的格式，可以说群雄并起。

先介绍几个正规军。

微软公司早就推出了AVI （Audio Video Interleave）文件格式。在MPEG4标准的基础上，后来又推出了ASF （Advanced Systems Format）。这是微软自己私有的格式，不公开。微软公司的播放器就是Windows Media Player。

苹果公司的QuickTime是一开始支持.mov 文件格式。后来转到标准的MPEG4格式上，支持.mp4文件，这也是比较先进的一种文件格式。

ADOBE公司的flash曾经非常红火，他们的.flv文件格式简单明了，也用得非常普及。但今天ADOBE已经放弃了对flash的支持，所以.flv格式也日益式微。

这么多文件格式有没有把大家搞晕？其实这些私家标准都差不多，都是定义了一些封装格式，就像一个箩筐，里面装的还是压缩过的音视频。

这里跳开去讲一个子贡赎人的故事。战国的鲁国，当时有一条法律：任何人只要在国外赎回沦为奴隶的鲁国人，鲁国政府会报销赎人的花销。子贡一次赎了个鲁国人，但是没有要国家报销！孔子知道后，痛心疾首，说：子贡你不对啊！从此以后再没有人愿意救赎身为奴隶的鲁国人啦！果然大家听说这件事以后，都不愿意再积极的救赎奴隶了，因为做好事救奴隶这件事，已经被子贡这个富二代拔高到别人没办法效仿的高度了。子贡给社会树立了一个极高的道德标准，自以为可以换来众人的效仿，没想到正是因为他这个标准，使得原本的善事，对于一般人来说成为了无力承担的行为了。

计算机行业也有这么个子贡就是当年的老大IBM。老大率先免费开放了PC机的技术标准，后来计算机界也就有了开放标准的传统。想要在计算机领域靠一个专利标准赚钱，比较困难。免费系统如Linux，开源软件在计算机界这么兴旺是有源头的。从传统电影电视工业来到互联网视频，大家不由感慨：贵圈真乱，收点专利费真难。

互联网视频的发展不能只是几家正规军啊。从DVD的加密被破解开始，把DVD拷贝成文件再分发就成了满足广大群众日益增长的文化需要的主要途径。分发视频需要一个格式，这样各路诸侯就出现了。

2000年，法籍影音发烧友罗达Rota （又有个名字叫Gej）就从微软把视频压缩技术偷了出来，成立了一家叫DivX 的公司发起一个完全开放源码的项目，名为"ProjectMayo"，他们的目标是开发一套全新的、开放源码的MPEG4编码软件。由于它完全符合ISO MPEG4标准，又是完全开放源代码，吸引了很多软件、视频高手参与。人力、物力大大投入，OpenDivX编码器和解码器原型很快便公布于众，之后又开发出具有更高性能的编码器Encore 2等等。高举开源共享的大旗，这是DivX的辉煌时期。

A person sitting on a table

Description automatically generated

图表 4 Rota， 钉在历史的耻辱柱上

可惜人性总是贪婪的。这家伙利用当时有限开源协议的漏洞(LGPL)，在革命胜利的前夕，封闭了代码，将DivX变成了公司私有财产。他们后来还推出了DIVX格式的DVD，进军DVD出租市场。

愤怒的开发者们被DivX利用了，他们也没有罢休，成立了一个新的真正开源的项目叫Xvid,继续他们的开源事业。Xvid就是DivX的名字倒过来，另外在英语中 “X”也往往表示“Fuck”的意思，从这个名字我们就可以看出被欺骗利用了的程序猿的怒火。

背信弃义的DivX的出租DVD业务发展也不顺利，一是他们的DVD必须要专有的机器才能放，二来蓬勃发展的互联网视频很快就取代了传统的音像出租业务。DivX在2006年上市，短暂的辉煌之后，2010年被收购，消失在历史的浪潮之中。

另一个大家可能有记忆的公司就是Realtime，他们开发的.rmvb格式体积小，也曾经风行一时。但随着网络带宽的增加，优势也渐渐消失了。

最后一统天下的还是要正规军，这就是当前最流行的H264视频压缩标准。

除了MPEG以外，世界上还有另一批专家在默默地研究视频压缩。ITU-T，全称International Telecommunications Union - Telecommunication Standardization Sector，即国际电信联盟——电信标准分局。该组织下设的VECG(Video Coding Experts Group)主要负责面向实时通信领域的标准制定。他们一直专注于研究可视电话中的视频压缩。他们的标准以H开头，从H.261, H.262 就和MPEG2一致，H.263也放进了MPEG4第二部分当中。

2003年，VECG 和MPEG又做了很多改进，携手推出了H.264标准。前面说过，MPEG4标准可扩展性比较好，所以H.264就做为MPEG4的第十部分，也称为AVC(Advanced Video Coding)，先进视频编码。

H.264的压缩率达到MPEG2的两倍以上，是H.263的1.5~2倍。一般可以轻松做到对原始视频压缩300~400倍。对比MPEG2，H264用了多种运动预测手段和图块大小。举个简单的例子，前面我们比过两只猫头部有变动，必须把猫的头重新编码发过去。更精细的比较可以找到猫头部也有不动的部分，比如猫耳朵只是移动了一点位置，但轮廓形状还是一样的。这样更精细的比较可以更好地消除掉冗余的信息。

但这样一来“消消看”的工作量就大多了。曾经有个判断视频压缩芯片性能的办法就是看它散热片的大小，压缩做得好的芯片必然计算量大，工作发热多，散热片就必须大。当然这只是初略估计性能，不排除有些发热多性能也差的垃圾设计。

2013年VECG和MPEG又联合推出了H,265也叫HEVC（ high-efficiency video coding），高效视频编码， “消消看 ”的本领更强，图块的尺寸定义更灵活，图块允许选择的角度更多，比H.264性能又有了大约50%的提升，也就是对原始视频可以做到500~600倍的压缩。

正规军可以融合当前的最新压缩技术，各个公司都竞相把自己的强项贡献出来，希望挤进世界标准，这样可以取得话语权，将来收取专利使用费用。

H.265在专利管理方面有些问题，专利池从H.264的一个变成了三个，造成使用者不方便，所以影响了它的普及。

# 电视行业

传统上电视行业是主要的视频传播媒介。数字化之后，电视行业从原来的无线广播发展出数字电视广播（包括卫星数字电视），有线电视。有线电视又经历了IP化改造，产业投入巨大。

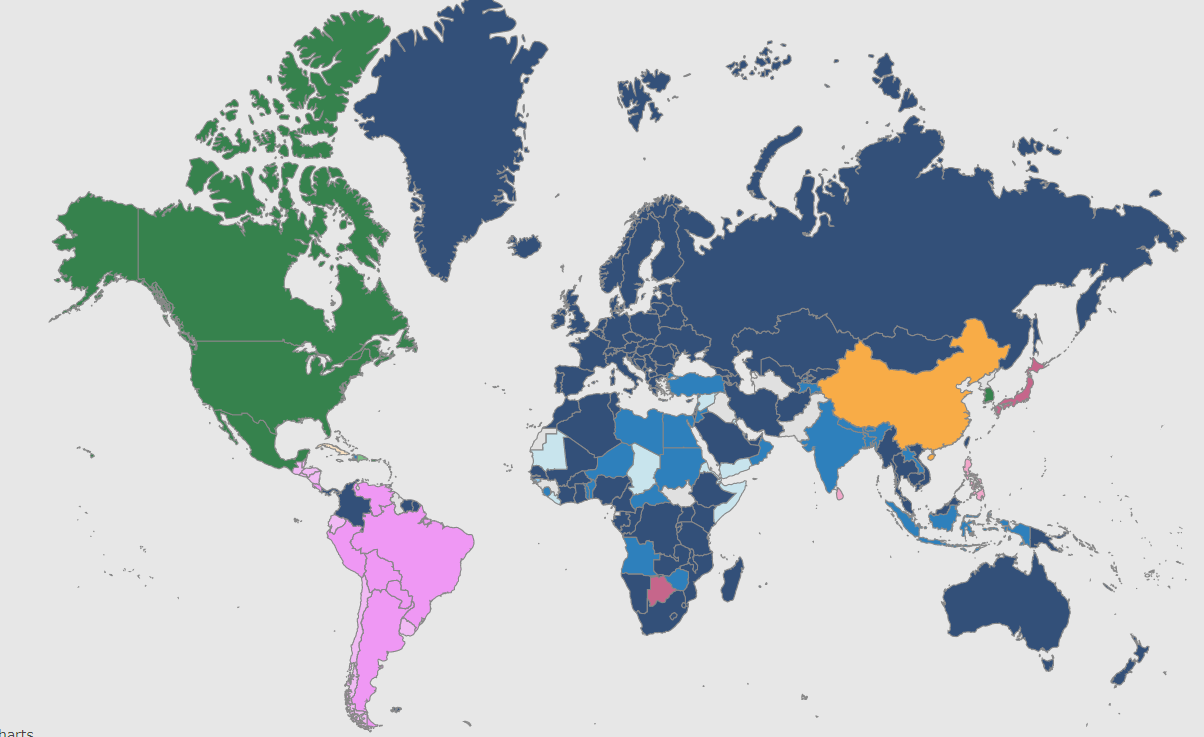
视频压缩标准的变动，体现在电视行业就是机顶盒的经常更新。这些成本不太可能从增加广告来回收，因为会影响观众体验，所以只能提高月租费用或者寄希望于附加服务比如电影点播等等。近年来OTT电视公司迅速兴起，利用互联网设备，彻底颠覆了传统电视公司。

OTT是Over The Top的缩写，来源于篮球等体育运动，是“过顶传球”之意，指的是篮球运动员在他们头之上来回传送而达到目的地。现在指通过互联网向用户提供各种应用服务。这种服务由运营商之外的第三方提供，不少服务商直接面向用户提供服务和计费，使运营商沦为单纯的“传输管道”。

Netflix, YouTube，乐视这样的公司都动了传统广播电视公司的蛋糕。由于他们没有前期大量基础设备的投入，在运营上没有地域的限制，又给予用户点播的自主权，发展十分迅速。

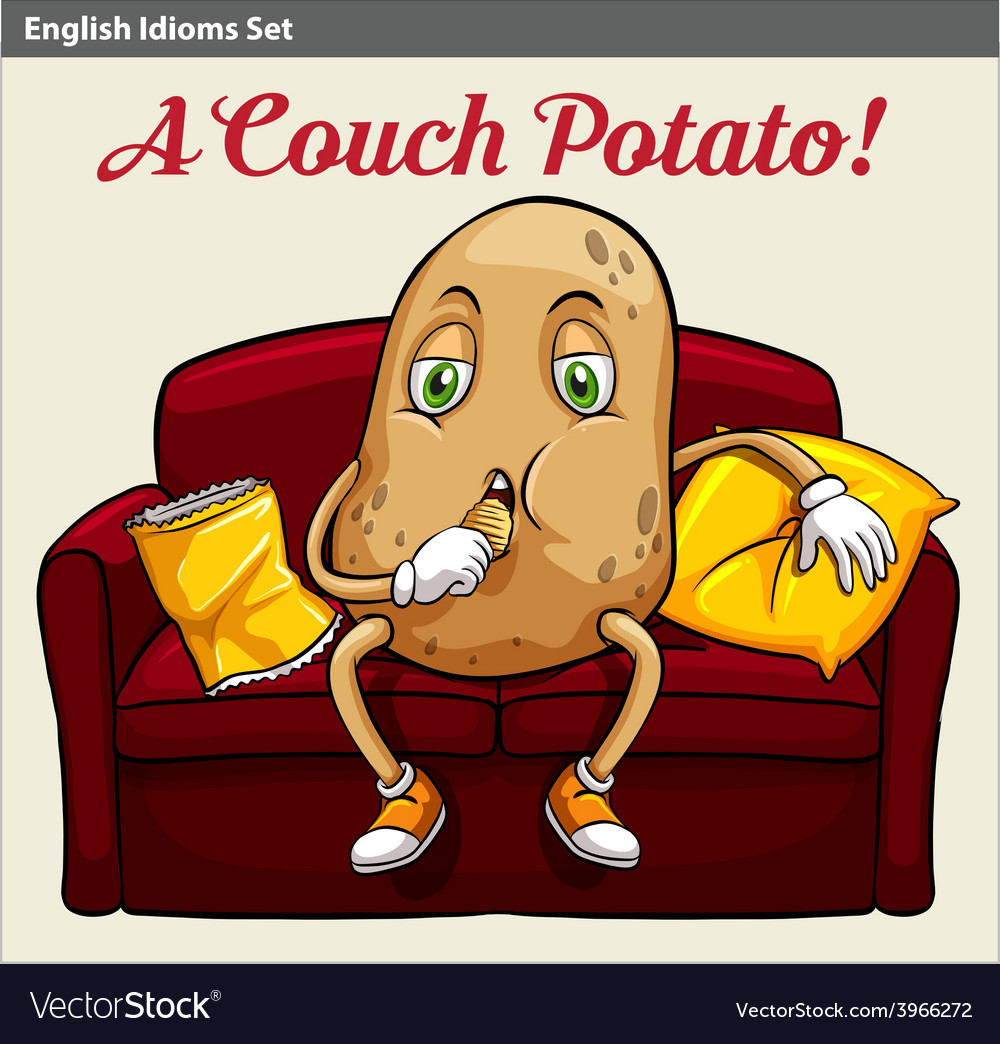
曾几何时，在数字电视领域我们可能看到如同彩色制式的NTSC，PAL，SECAM那样的大战并没有上演多久，地面数字电视的几大制式：美国的ATSC，欧洲的DVB-T，日本的ISDB-T，中国的DTMB最终也是因为广播电视不再是观众的主要视频来源而归于沉寂。

熟悉中国的人可能对DTMB经历的清华的多载波标准和交大的单载波标准的大战记忆尤新。最后忙于这些标准的公司可谓是竹篮打水一场空。观众跑了比啥都问题严重。



图表 5 地面数字电视广播标准，中国DTMB仅有中国内地及港澳使用。

观众再也不愿意当Couch Potato了（Couch Potato的意思是坐在沙发上手拿遥控器傻等节目的土豆），他们要求自己选择能看什么，OTT正好符合他们的要求。



图表 6 沙发上的土豆

据统计，今天70%~80%的网络流量是视频流产生的，网络电视发展迅猛。

Blockbuster, Circuit City这些DVD出租商店，卖碟小贩慢慢也消失在历史长河之中，由于人们能很容易得到想要的视频信息，足不出户的宅男宅女开始出现。

# 新的时代

互联网发展迅速，带宽日益增加，视频也从高清向超高清(4K)，及8K发展。

另一方面观众不满足于观看，他们也想成为视频的发布者。2019年，YouTube平均每天就有72万小时的视频上传，一些节目的观看人次超过了10亿，一些热门频道的年收入超过了千万美元，不亚于一个小型电视台的年收入。

直播的兴起在游戏，体育，新闻中有广泛的应用。可以说每个直播者就是一个小小的电视台，他们可以有过百万，千万的实时观众。视频不再是流向观众而是反过来从观众流向其他观众，视频传输从原来的单向变成多向。

网络云游戏的发展又对视频压缩的实时性提出了更高的要求。

虚拟现实技术同时需要实时性和高分辨率。新的3D视频，沉浸式视频的标准也在制定当中。

新时代的活雷锋公司google针对H264,H265费用问题，在制定他们的VP9/VP10视频压缩标准之后推出了免费的AV1压缩标准，比H.265效率更高，虽然目前计算的复杂度还太高，但免费是一个多漂亮的词啊。

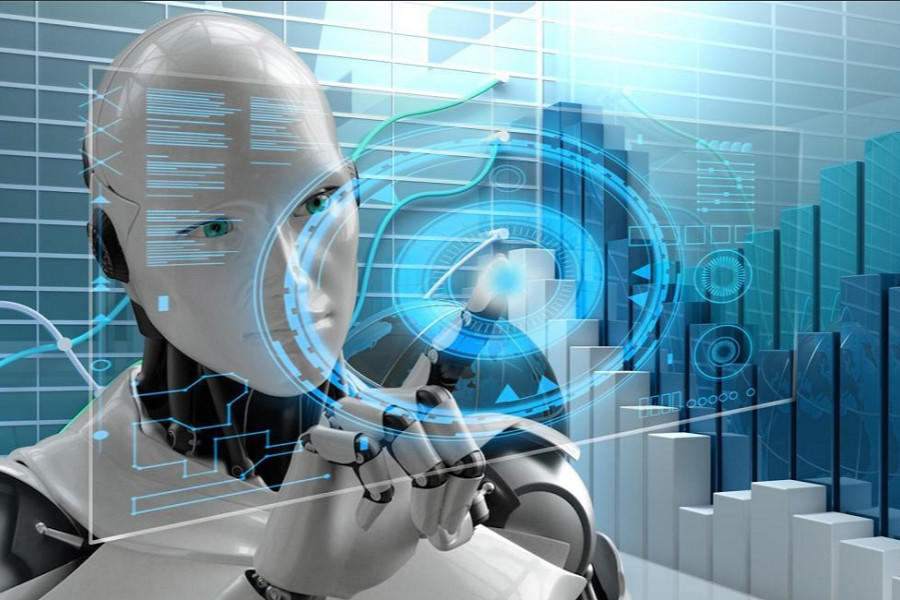
中国一向重视自主知识产权，独自开发了同H.264媲美的AVS1标准，以及和H.265媲美的AVS2标准，目前瞄准压缩千倍的AVS3标准。

正规军VECG 和MPEG还是在网罗最好的技术，准备放在下一代的H.266当中，这个标准也叫VVC (Versatile Video Coding)灵活视频编码技术，这个技术提供了分层的概念，能根据带宽提高不同质量的视频。这些新一代的视频压缩的目标是将原始视频压缩千倍。

视频压缩技术还有没有新的突破呢？人工智能的发展为视频压缩提供了一些新的方向。比如我们要压缩一个猫的视频，我们也许只要简化成：一只成年波斯猫，白色，胖，坐着，正前方斜上观察角度，距离5米，再加上波斯猫的标准模型库就可以还原出高清晰的视频。人类脑记忆容量有限，但我们对视觉的记忆可以很多，比如可以记得童年的画面，这些都不是根据图像像素记忆的，而是根据概念比如“猫”，“房子“，“树”等等，这样的压缩率是惊人高效的。

人工智能视频压缩还有一个好处，就是恢复出的视频质量甚至能超过输入的视频。比如一只拍摄得模模糊糊的猫可以通过预先存储的高质量猫的图像，恢复出一只清晰的猫。这只猫跑到光照不好的地方，我们也能把它的图像完美地根据前面在光线好地地方拍摄的图像恢复过来。

最近，对视频的要求出现了一些新的有趣的变化，观众不再是人了，而是机器人！大部分视频是被机器人在检查。比如自动驾驶，交通违规，犯罪嫌疑人识别，人脸识别等等。直播的兴起又要求能够实时鉴别黄色，政治不正确的内容，大量的视频靠人24小时不休息去检测是不现实的，只能通过机器人(人工智能)来识别。



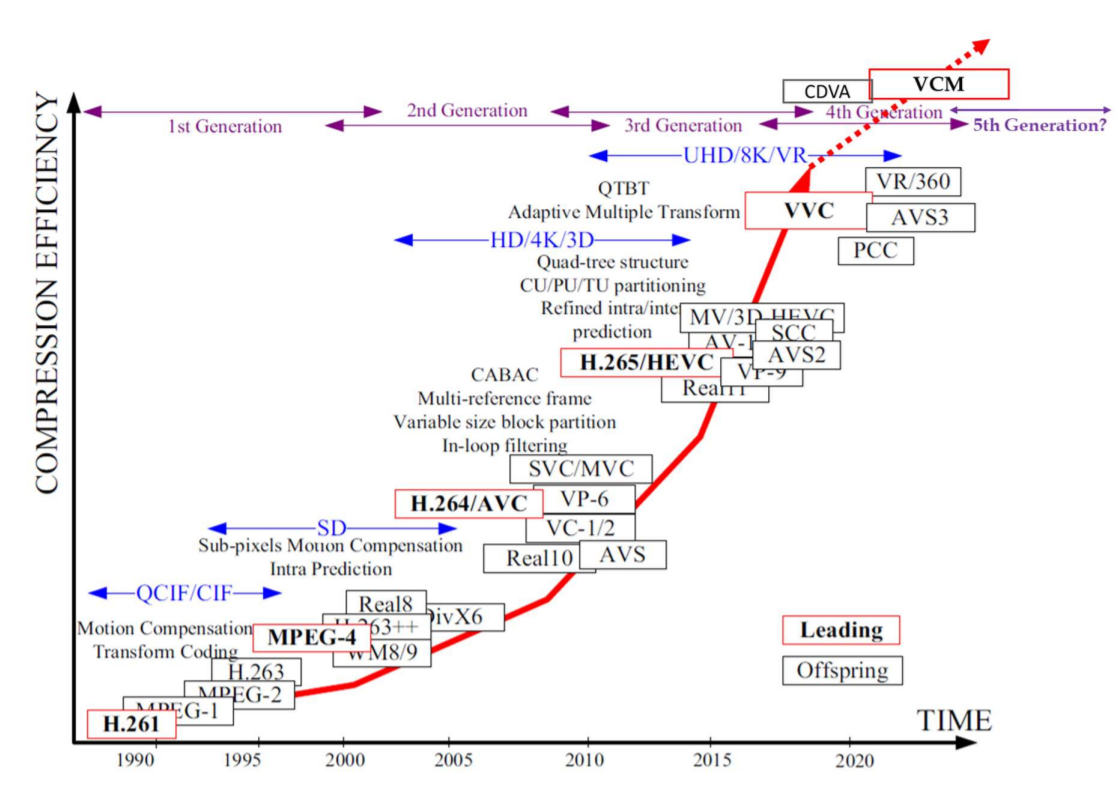
图表 7 机器视觉

机器人看东西和人看东西当然不同。一些人眼重视的东西，机器人可能不重视。反之，人不重视的东西，机器人可能非常需要看清，比如车牌号码等等。

现在VCM (Video Coding for Machine) 视频机器编码正变得火热，MPEG-VCM专家组正准备推出新的标准。

雄关漫道真如铁，而今迈步从头越。

这张图简单地归纳了一下视频压缩的历史和未来。从第一代的MPEG1,MPEG2, 到第二代的H.264,再到第三代的H.265，到第四代的VVC(H.266)再到未来的VCM。压缩效率越来越高，支持的图像越来越清晰。



图表 8 视频压缩标准的演进 (gyrfalcon 公司)

从视频技术这些年的发展来看，新的需求，新的模式，新的技术层出不穷。视频作为信息的重要形式是信息革命的一个重要方向，希望更多的程序猿，攻城狮们能投身于这个行业。