

# 智能时代 视频编解码技术的发展及行业应用

■ 文 / 李浙伟 浙江大华技术股份有限公司

## 一、背景介绍

人类获取现实世界的信息 80% 以上通过视觉，从机器理解视频的角度，视频通过扫描形成连续的帧，一帧对应一幅图像，图像帧以一定的速率播放形成连续的视频。而连续帧之间的高相似度导致了大量的信息重复冗余，因此我们可以通过视频编解码技术用尽量少的数字信息来表示原始的视频，并在解码端恢复视频图像后，从人眼的视觉暂留效应角度来看，尽量接近原始图像。

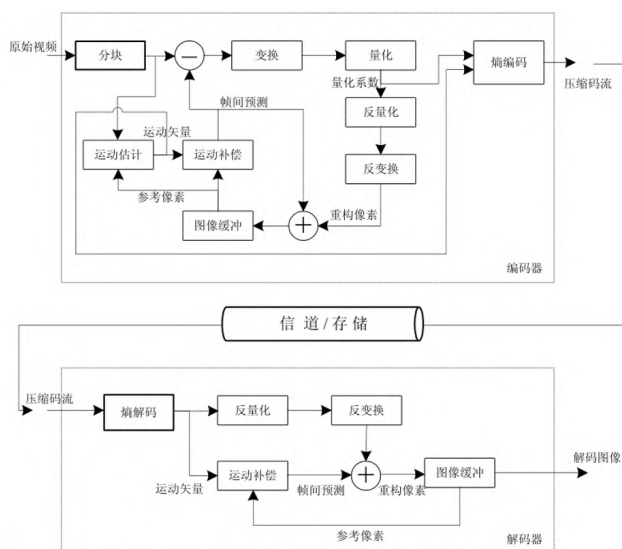


图1 视频编解码器基本框架  
(图片来源论文：高性能视频编解码芯片结构设计  
张鹏 中国科学院研究生院 2009)

基于现有的视频编解码框架，编解码标准的更新和优化主要围绕各个技术子模块流程开展创新，编解码技术主要围绕：混合精度量化、帧内 / 帧间的预测、变换 / 量化 / 反量化 / 反变换、环路滤波等流程中的关键模块以及解码工具和解码算法开展自主提升。从近十年的专利角度看，熵编码、块划分、环路滤波技术及预测、变换量化、码率控制等传统视频编码技术的专利积累最多，在码率控制、率失真优化技术上呈现逐步上升趋势。从技术原理角度看，视频编解码主要围绕

帧采样波形（空间及时间的相关性）和帧内容差异（形状及运动的差异性）两个方向开展研究，常见的有预测编码和变换编码，以及融合两种方案的混合编码。

## 二、视频编解码技术标准的发展历程

国际上制定视频编解码标准的组织主要是ITU-T和ISO/IEC。ITU-T主要制定如H.261、H.262、H.263、H.264标准，面向实时视频通信应用；ISO/IEC主要制定MPEG-1、MPEG-2、MPEG-4标准，面向视频存储、广播视频和视频流应用。同时两个组织ITU&MPEG通过进一步合作输出H.262/MPEG-2、H.264/AVC、H.265/HEVC、H.266/VVC等标准。

国内制定视频编解码标准的组织主要是AVS和SVAC。

**1.AVS：**《信息技术先进音视频编码》由我国数字音视频编解码技术标准工作组自主牵头，基于技术先进性，法律法规和专利管理完备性，制定的具备自主知识产权的第二代音视频编码标准，主要围绕音频、视频、系统、数字版权管理等四个方面

展开深度的技术研究。AVS编码效率与H.264相当，其硬件实现复杂度低，采用混合编码框架，技术方案达到业界领先水平，当前AVS编解码标准已发展到AVS3，同时AVS4也在规划中，AVS提出的技术优化创新以及对产业化的推动得到国际标准组织的认可。

**2.SVAC：**《安全防范监控数字视音频编解码技术标准》由我国公安部第一研究所牵头，在现有视频编码标准技术的基础上，针对安全防范特定的应用场景形成了一套我国自主可控的新标准。主要创新点基于自主可控的加密特性，支持安全防范专用的音视频内容的有效管理，支持加密和认证，保证安全防范数据的保密性、真实性和完整性，同时支持高精度视频数据编码和感兴趣区域（ROI）可变速率编码。该标准因为开放与兼容方面的影响，与市场上主流的视频系列化设备不能实现很好的对接，由此限制了市场大规模的应用。

互联网领域本身存在巨量的视频分享需求，而HEVC等标准又面临昂贵的使用费，因此国际互联网各大企业联合成立开源媒体联盟，推出了新的视频编解码标准，如VP9/AV1。

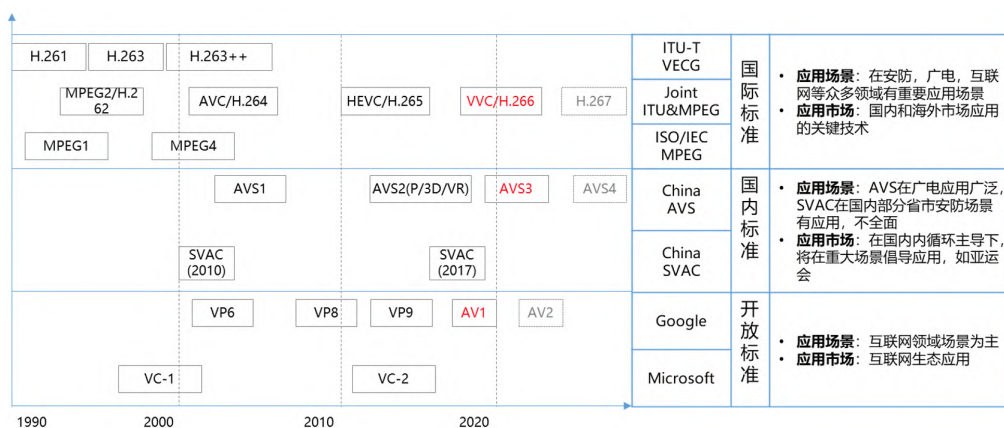


图2 各组织视频编码标准时间轴

### 三、安防行业对视频编解码技术的理解及在显示控制领域的发展趋势

视频采集的端侧设备类型越来越丰富，像素越来越高。摄像机的清晰度从 200 万到 800 万再到 3200 万像素，各类媒体播放及超高分视频服务器输出的视频源像素也已达 8K 甚至更高级别，众多显示屏厂商推出了 4K/8K 的超高清显示设备，安防行业逐步进入超高清时代，超高清的体验来自高清晰的图像传输及显示。因此视频编解码的应用不仅需要实现高效的视频压缩效率，同时需要兼顾高清画质与超低编解码及传输延迟，这是当下安防领域视频编解码重点提升和优化的方向。

在 1080p 视频时代，H.264 的视频几乎可以在任何设备上流畅运行，因为它有着很全面的硬件兼容能力，H.264 编解码的复杂性不高，但也导致了压缩视频的效率不高。除此之外，H.264 不支持 4K 以上分辨率的视频，面对越来越多的 3D 立体视频流，通常都是达到 1200 万像素以上分辨率的文件，H.264 无法有效支持。同时 H.264 的图像质量不高，原本黑色的画面看起来是灰色的，当比特率很低的时候，编码结果呈现块状，极大影响显示效果。面对传统 H.264 编解码的缺点，随着全域感知应用在行业的普及，新一代编解码技术在安防领域将围绕不同的场景持续发展，如在可伸缩编码技术、多视点编码技术，智能编码、高压缩性能方向不断创新，而在显示控制领域将转向高画质、低延迟方向转变。

#### 1. 可伸缩编码技术 (Scalable Video Coding)

可伸缩视频编码技术把视

频信号编码成一个基本层和多个增强层，分层的形式满足了不同带宽条件下的视频质量特性需要，本技术基于在时域、空域和质量等方面的灵活可调，以及只需编码一次的特性将推动其在安防行业多码流场景的快速应用。

#### 2. 多视点编码技术 (multiview video coding)

多视点视频编码技术是基于 3D 立体感知视频的编码技术，在安防行业，随着多目、多视角视频汇聚的全景拼接前端摄像机的大规模应用，多视角立体防控、多摄像头联动和交互控制等方案的普及与场景落地，多视点编码技术取得了越来越快的发展。

#### 3. 智能编码技术

智能 ROI 感兴趣区域的视频编码技术，根据对图像中感兴趣区域的不同等级进行量化参数控制，实现不同等级的码率分配，有效实现了编码码率和画面质量之间的平衡。

在此技术上亦衍生出针对各类场景的感知编码及优化策略。在静态场景下，合并处理整体未发生改变的大面积同色像素点；在动态场景下，针对相邻画面之间的变动情况，将画面中感兴趣区域和非感兴趣区域分离，采用不同的编码方式分开编码整合。背景模型在环境未变化的一定时间段内是固定不变的，只需提取而不需要重复编码，并通过只传

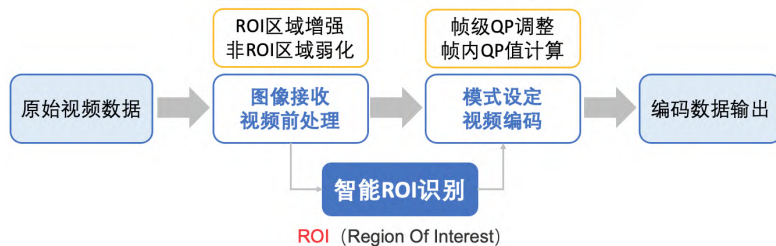


图3 智能编码技术



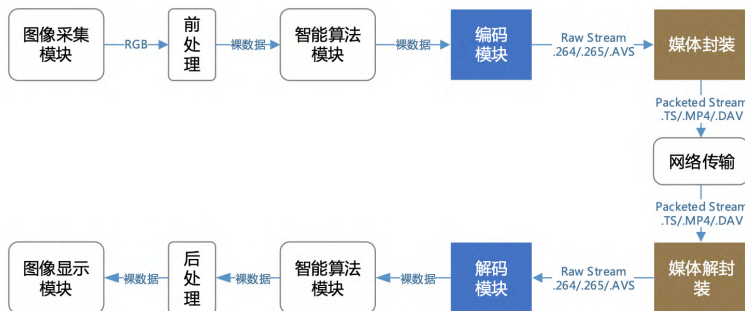


图4 视频图像业务系统编解码处理流程

输有变动的区域,智能图像算法极大地减少数据处理量,从而缩短编码处理时间。智能编码技术加入智能分析技术,采用先进的场景自适应码率控制算法可实现比 H.265/H.264 更高的编码效率,码率在白天可降低 30% 以上,夜晚可降低 70% 以上,同时提供高质量视频可显著降低存储和传输成本。

新一代智能编码主要面向视频的感知、分析和理解,基于特征的感知和基于语义的认知编码方法,我们可称之为感知编码和语义编码。感知编码的典型应用可表现在视频会议的视频画面中,传送每一位与会者的面部照片,只需对面部关键帧进行编码,对于非关键帧降低采样获取边缘信息,在解码端只需要输入边缘信息。通过使用 GAN 对抗网络生成视频,并用关键帧来做训练可大大地降低编码码率。随着各类数据驱动、计算芯片的算力提升,深度学习编码技术的快速突破,将带动感知和认知编码模型质的飞跃。

#### 4. 高压压缩编码技术

对于视频流编解码而言,在相对画质不变的情况下,压缩比、传输带宽、存储容量是项目建设的设计关键要素。在分辨率不断提升

的背景下,压缩性能的提升能够为网络和存储带来极大地开销节省,大幅减少基础设施投入,如最新的 VVC 视频编解码技术研究,其目标压缩性能将带来 50% 的提升。

#### 5. 高画质编码技术

安防系统视频图像处理业务

的一个常见流程如图 4 所示,由于人眼对亮度的敏感度远高于色度,因此人眼看到的有效信息主要来自于亮度,根据仿生的原理,计算机处理采集到的视频裸数据信号采用亮度信息和色度信息分别采集的方式。目前安防领域在处理视频、图片、相机等应用中常用的颜色空间为 YUV 格式。

YUV 格式用一个亮度 Y 分量和两个色度分量表示,分别称为 U (蓝色分量与亮度的色差) 和 V (红色分量与亮度的色差)。色度信号全采样的格式主要包括 YUV444、YUV422、YUV420 等。YUV444, Y 分量和 UV 分量一一对应; YUV422, 每两个 Y 共用一组 UV 分量; YUV420, 每 4 点 Y 分量对应 1 个 U 分量和 1 个 V 分量。从图像解析度角度看, YUV444 最优; 在

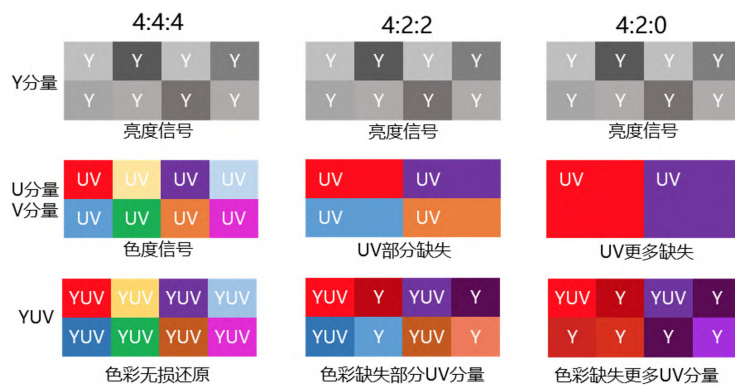


图5 YUV 颜色空间

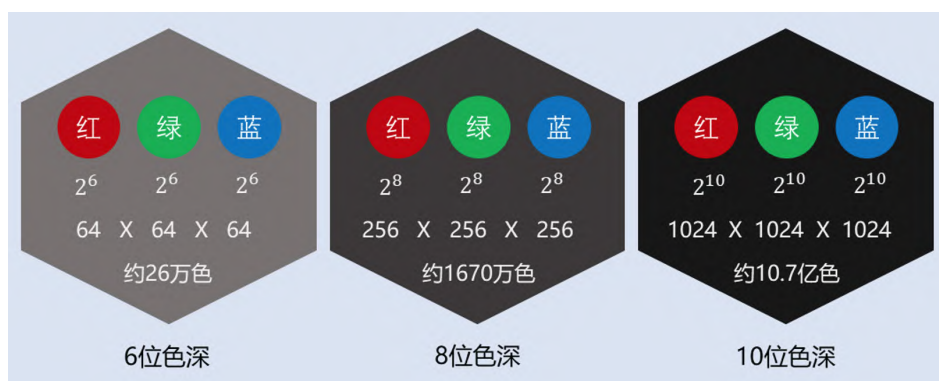


图6 不同位数色深换算

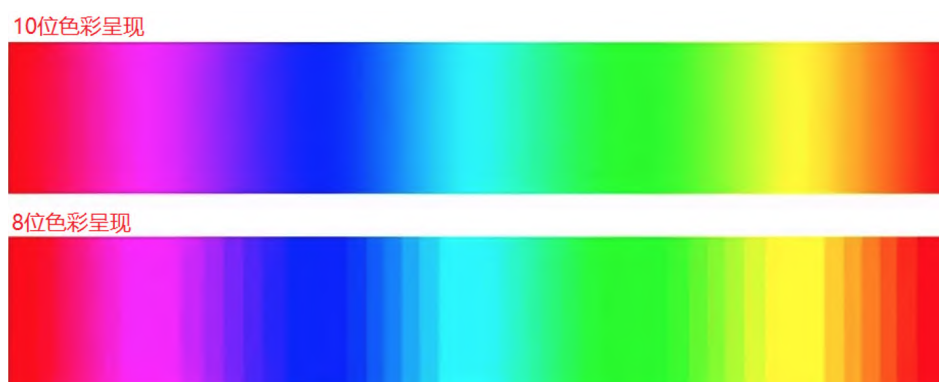


图7 不同色深显示画面对比

信号传输带宽上, YUV420 最节省。因此在普通的视频编解码算法上, 为节省传输带宽开普遍采用 YUV420 或者 YUV422 的采样格式, 但是为了达到所见即所得的显示效果, 真实还原图像色彩的场景, YUV444 真彩无损显示是更好的选择。

色深是色彩深度的简称, 通常表示存储一个像素的颜色所用的位数, 色彩深度越高, 可用的颜色就越丰富。色彩深度通常用“n 位颜色”来表示: 若色彩深度是 n 位, 则有 2 的 n 次方种颜色 (包括黑白) 选择。颜色为红绿蓝三原色, 所以 n 位色深可以显示的颜色为: 8bit 色深即 2 的 8 次方, 可显示 1677 万种颜色; 10bit 色深即 2 的 10 次方, 可显示 10.7 亿种颜色。所以 10bit 能显示的颜色差不多

是 8bit 的 64 倍。

最直接的视觉感受就是 8bit 色深的显示色彩视觉呈现带状分布, 而 10bit 色深的显示色彩间视觉呈现过渡更自然且柔滑细腻, 对于超高清视频的处理而言, 10bit 能更好地还原图片或视频图像。

## 7. 低延迟编码技术

在视频显示控制的应用中, 影响用户真实体验的端到端延迟主要包括设备内部延迟和网络传输延迟。以 1080p60 视频传输为例, 常见的端到端延迟 60ms–80ms 时, 鼠标操作时就会明显感受到迟滞感, 因此视频编解码设备端全链路的延迟缩小就成了优化高清视频流畅交互体验的关键。对于安防领域的视频编解码系统而言, 端到端的延迟主要集中在

在采集编码、网络传输、解码显示。

为了进一步降低端到端的视频编解码延迟, 大华新一代编解码技术基于视频图像多切片 (slice) 并行处理的方式来实现延迟的进一步缩小, 通过多切片并行处理的方式, 以及针对网络传输的特定优化, 端到端的视频编解码延迟可以降低到 17ms。

(1) 视频采集模块首先对采集到的完整画面进行拆分, 拆分为多个切片模式。

(2) 当切片 1 采集完成就可通过编码单元进行编码, 完成之后通过网传, 解码并更新显存显示对应的是切片 2, 切片 3 不需要更新显存可降低显示等待时间。

(3) 拆分为多个切片之后, 每个切片由编码单元编码输出, 每个切片的编码过程相互独立, 并行对多个切片进行编码, 加快编码速度。

(4) 编码过程中切片内部可参考的信息减少, 同时每个切片都包含专有信息, 单个切片组装的码流更小, 降低了网络波动影响, 有利于降低整体延迟。

(5) 多切片模式需要解码端同步配合实施保障机制, 单个切片所需要处理的范围较小, 有效减少解码开销。

(6) 多切片异步处理, 当某个切片发生异常, 错误仅限于单个切片内部不对外扩散, 在解码及全局显示时出错影响范围大幅降低。

## 四、编解码器与视频综合平台

编解码应用的本质就是将源端视频信号进行连续采集、编码、压缩, 经网络传输后, 再解码还原成画面的过程, 在安防行业的各种显示控制应用领域, 如视频管理中心、指挥中心场景的使用过程中, 常见的困难及痛点有:

**一是编码画质不高带来的色彩还原不足。**以

Ycbcr444 的图像为例, 传统的编码芯片只能支持到 YUV4:2:0 的编码, 相当于每个像素点等价只有 12bit 的色深。这样会造成单像素点颜色采样不足, 同时像素点之间的共用及相互参考, 造成边缘不清晰 (如字体的颜色晕染), 并且最终采集到的画面和实际上墙的画面之间存在色差。

**二是编解码延迟高带来的交互控制体验无法满足实时应用。**50ms 以上的端到端编解码延迟, 人

眼能够明显的感受到迟滞 (如挥手对比延迟, 鼠标操作迟滞), 同时当视频源的采集帧率低时, 高速动态的画面显示会呈现顿挫。

**三是传统集中式设备无法快速根据实际的规模化多点应用灵活扩展。**如在指挥控制场景, 各个部门的不同高清视频信号的汇聚及调度, 面临信号格式多样, 信号集中布线纷繁复杂, 无法支持多点协同操作。

**四是各类视频信号无法快速有效联动。**如何加快解决应急管理中有有效信息获取、联合分析效率提升、精准指挥效率等问题。

大华基于在 H.265 芯片定义和智能编码技术上的积累, 主要采用 H.264/H.265 编解码标准技术以及在此技术上自主深度优化的 SmartH265 智能编码技术, 以及在采集编码、传输、存储、解码显示及控制领域自身独到的理解, 发布系列化的视频编解码器产品、解码器产品以及编解码节点融合产品, 下文对编解码产品在视频管理中心和指挥中心领域的应用进行详细阐述。

### 1. 管理中心应用

管理中心的编解码产品基于当前业务的系统集约化、分辨率高清化、空间占用小型化、操作简便

化等需求应运而生。特别是在目前编解码芯片性能翻倍提升、产品设计进一步高密度化、高端显控技术持续下沉的背景下,解码器产品逐步承担了管理中心场景里的重要角色。其具有多协议接入、解码兼容性强、超高清采集、超高清解码、超高清输出等优点。

## 2. 指挥中心应用

指挥中心做为应急指挥的神经中枢,具备处理突发事件以及为决策提供支持服务的职能。因此系统响应时间要求、视频实时性、视频真实性至关重要,同时指挥中心需融合调度各层级的子系统,传统集中式视频综合平台快速拓展存在不足。在大华新一代指挥中心解决方案的场景应用中,核心关键的输入节点、输出节点和拼接节点之间依托高画质、低延迟的编解码技术以及去中心化架构分布式的部署方式,实现各类信号接入,为指挥中心建设提供有力保障。

## 3. 强大的信号处理和图形显示能力

对各种 Video、RGB、DVI、HDMI、SDI 信号源的图形具有相同的显示处理能力,所有信号实时显示。实现单屏显示、跨屏显示、叠加显示、整屏漫游等多样化的显示模式,单屏开设多个图像窗口,整屏可以组合成多图层画面布局。输入分辨率支持不少于  $3840 \times 2160@60\text{Hz}$  像素,输出分辨率支持不少于  $3840 \times 2160@60\text{Hz}$  像素。从信号采集编码至解码显示输出延时低至 17ms,为交互控制提供低延迟基础。全链路基于高画质的编码视频传输,强大的解码性能实现超高清上墙还原真实影像。

**(1) 人机分离:** 实现“人机分离”的运行环境,用户可以选择将所有服务器统一放置在机房,统一

部署开展管控,减少重要数据丢失风险,提高运维便捷高效,保障坐席人员业务操作上的专注,从而提升指挥实战的工作效率。

**(2) 一屏多系统:** KVM 输出节点具备单台显示器同时查看并跨屏操控不少于 64 个不同业务系统的画面功能,单屏幕分割模式下支持鼠标滑屏,任意一个分割区域都可通过鼠标滑动对相应的主机进行 KVM 控制,实现计算机内容(文档编辑、PPT 放映)的交互控制。

**(3) 一人多业务:** 单人可通过一套键盘鼠标、多个显示终端完成对多业务系统信号源的操作,实现多业务平台、多系统之间的实时切换,实现跨网段、跨系统极速编辑操作,开展高效协作,数据实时调取,快速的推送与分发形成有效联动,支撑各类事件处置的分析决策,从而在整体上提高应急响应能力。

## 五、结束语

随着视频编解码技术的发展,视频显示分辨率步入 8k 时代,未来甚至到 16k,在保证图像质量的情况下,传输带宽和视频存储空间之间始终存在平衡需求,随着硬件层面计算芯片技术和软件层面深度学习技术发展,基于人工智能深度学习算法的跨界,神经网络智能编码技术的探求,我们相信中国在技术、专利、标准、产品、应用五个环节的自主创新上能够持续推进自主产业链的形成,实现跨越式发展,并在保障信息安全保密的要求下实现互联互通开放性。面向 AIoT 全场景的各领域,基于视频图像的智能化应用如火如荼发展,编解码技术的发展不仅需要满足越来越高的用户体验需求,同时也会按照行业场景应用实现自主创新。✎