

【H.264/AVC视频编解码技术详解】一. 视频信息与压缩编码_网络_Workshop of Wenjie.Yin-CSDN博客

 blog.csdn.net/shagoneal/article/details/52080975

《H.264/AVC视频编解码技术详解》视频教程已经在“CSDN学院”上线，视频中详述了H.264的背景、标准协议和实现，并通过一个实战工程的形式对H.264的标准进行解析和实现，欢迎观看！

“纸上得来终觉浅，绝知此事要躬行”，只有自己按照标准文档以代码的形式操作一遍，才能对视频压缩编码标准的思想和方法有足够深刻的理解和体会！

链接地址：[H.264/AVC视频编解码技术详解](#)

本节视频免费

一、人与世界的交互

从远古时代人类文明诞生之日起，人类就在不断为适应环境、改造环境而艰苦奋斗。其中最为基础的前提就是利用感官对外界信息的获取。利用各种感官，人类可以与世界环境进行多种不同的信息交互，例如：

- 嗅觉：识别各种气味，辨别环境变化和食物、饮水的质量等；
- 听觉：识别同类的联络信息和天敌等危险信号等
- 味觉：挑选最适宜的食物
- 触觉：制作与使用工具时非常重要

另外，最重要的一种自然就是视觉。据统计，在人的各种感官中，视觉占据了超过7成的信息获取量。而且视觉可以使人对环境变化做出最直接的反映。

在文明的发展过程中，人们不满足于仅仅靠着口述记录所看到的影像，而是希望能用更加直观的形式将其记录下来。经过了多年的发展，经过多年的发展，视频已经成为记录和重现信息最为高效的方式，可以在相对很短的时间内传递大量的信息。

- 视频通过其中每一帧的图像表达信息；
- 视频包含的音频可提供大量信息；
- 视频通过图像的运动、场景的变换提供信息；

综上所述可知，视频信息提供了最为接近人的直接体验的信息表示方式。

二、视频信号的表示方法：RGB与YUV

真实世界中的影像与早期的视频处理与传输系统所处理的都是模拟信号。然而为了能适应现代的计算机、网络传输与数字视频处理系统，模拟的视频信号必须转换成数字格式。

在数字格式的视频信号中，组成视频的基本结构是一帧帧连续而相关联的图像。组成每一帧图像的基本结构是成平面紧密排列的像素，每个像素代表图像中的一个彩色的点。为了使得像素以彩色表示，每一个像素由颜色的三基色取不同的分量构成：

- R:Red——红色分量；
- G:Green——绿色分量；
- B:Blue——蓝色分量；

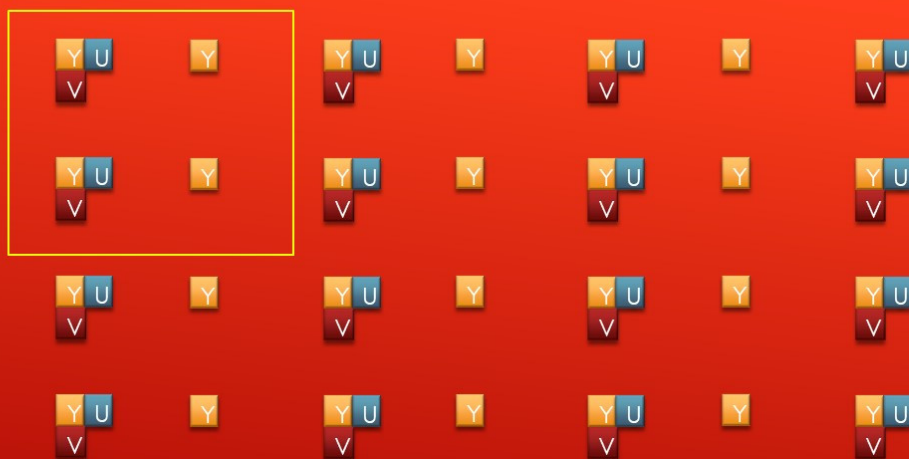
使用这种方式表示彩色图像的方式为RGB颜色空间。RGB颜色空间常用于显示器系统。通过这种形式表示的图像，每个像素的每一个颜色分量用1个字节表示，则可以表示 $256 \times 256 \times 256$ 种不同的颜色。在常见的图像格式中，如位图（bmp）格式以RGB形式保存数据。

在实际的编解码等视频处理的过程中，YUV格式比RGB格式更为常用。在YUV格式中，一个像素由亮度分量和色度分量表示，每一个像素由一个亮度分量Y和两个色度分量U/V组成。亮度分量可以与色度分量一一对应，也可以对色度分量进行采样，即色度分量的总量少于亮度分量。

在YUV中之所以采用这样的方式，主要是因为人的感官对亮度信息的敏感度远高于对色度信息。因此相对于其他像素格式，YUV的最大优势是可以适当降低色度分量的采样率，并保证不对图像造成太大影响。而且，使用这种方式还可以兼容黑白和彩色显示设备。对于黑白显示设备，只需要去除色度分量，只显示亮度分量即可。

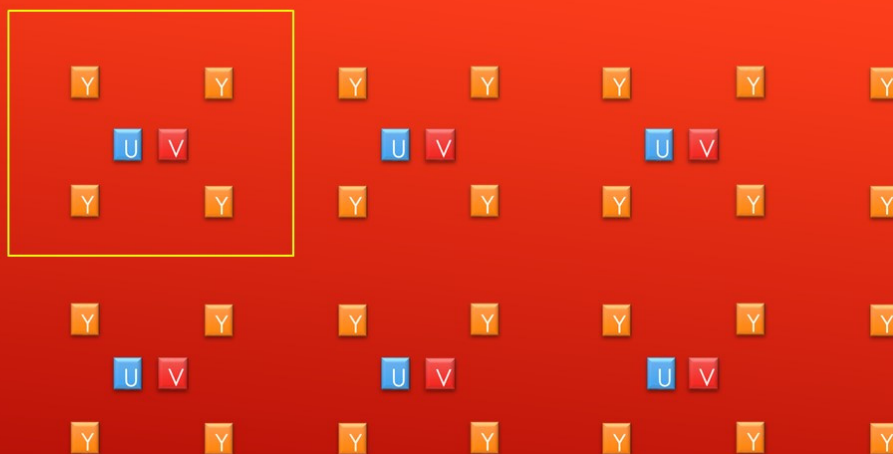
在YUV中常见的色度采样方式有4:4:4、4:2:2和4:2:0等，如下图所示：





YUV 4: 2: 2格式

<https://blog.csdn.net/shaqoneal>



YUV 4: 2: 0格式

<https://blog.csdn.net/shaqoneal>

三、视频压缩编码

编码这一概念在通信与信息处理领域中广泛使用，其基本原理是将信息按照一定规则使用某种形式的码流表示与传输。常用的需要编码的信息主要有：文字、语音、视频和控制信息等。

1. 为什么需要对视频编码

对于视频数据而言，视频编码的最主要目的是数据压缩。这是因为动态图像的像素形式表示数据量极为巨大，存储空间和传输带宽完全无法满足保存和传输的需求。例如，图像的每个像素的三个颜色分量RGB各需要一个字节表示，那么每一个像素至少需要3字节，分辨率1280×720的图像的大小为2.76M字节。

如果对于同样分辨率的视频，如果帧率为25帧/秒，那么传输所需的码率将达到553Mb/s！如果对于更高清的视频，如1080P、4k、8k视频，其传输码率更是惊人。这样的数据量，无论是存储还是传输都无法承受。因此，对视频数据进行压缩称为了必然之选。

2. 视频信息为什么可以被压缩

视频信息之所以存在大量可以被压缩的空间，是因为其中本身就存在大量的数据冗余。其主要类型有：

1. 时间冗余：视频相邻的两帧之间内容相似，存在运动关系
2. 空间冗余：视频的某一帧内部的相邻像素存在相似性
3. 编码冗余：视频中不同数据出现的概率不同
4. 视觉冗余：观众的视觉系统对视频中不同的部分敏感度不同

针对这些不同类型的冗余信息，在各种视频编码的标准算法中都有不同的技术专门应对，以通过不同的角度提高压缩的比率。

3. 视频编码标准化组织

从事视频编码算法的标准化组织主要有两个，ITU-T和ISO。

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，主要制定了H.261/H263/H263+/H263++等标准。

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)，即移动图像专家组主要负责面向视频存储、广播电视、网络传输的视频标准，主要制定了MPEG-1/MPEG-4等。

实际上，真正在业界产生较强影响力的标准均是由两个组织合作产生的。比如MPEG-2、H.264/AVC和H.265/HEVC等。

不同标准组织制定的视频编码标准的发展如下图所示：

□

四、视频压缩编码的基本技术

为了专门处理视频信息中的多种冗余，视频压缩编码采用了多种技术来提高视频的压缩比率。其中常见的有预测编码、变换编码和熵编码等。

1. 预测编码

预测编码可以用于处理视频中的时间和空间域的冗余。视频处理中的预测编码主要分为两大类：帧内预测和帧间预测。

- **帧内预测**：预测值与实际值位于同一帧内，用于消除图像的空间冗余；帧内预测的特点是压缩率相对较低，然而可以独立解码，不依赖其他帧的数据；通常视频中的关键帧都采用帧内预测。
- **帧间预测**：帧间预测的实际值位于当前帧，预测值位于参考帧，用于消除图像的时间冗余；帧间预测的压缩率高于帧内预测，然而不能独立解码，必须在获取参考帧数据之后才能重建当前帧。

通常在视频码流中，I帧全部使用帧内编码，P帧/B帧中的数据可能使用帧内或者帧间编码。

2. 变换编码

目前主流的视频编码算法均属于有损编码，通过对视频造成有限而可以容忍的损失，获取相对更高的编码效率。而造成信息损失的部分即在于变换量化这一部分。在进行量化之前，**首先需要将图像信息从空间域通过变换编码变换至频域，并计算其变换系数供后续的编码。**

在视频编码算法中通常使用正交变换进行变换编码，常用的正交变换方法有：**离散余弦变换(DCT)、离散正弦变换(DST)、K-L变换等。**

3. 熵编码

视频编码中的熵编码方法主要用于消除视频信息中的统计冗余。由于信源中每一个符号出现的概率并不一致，这就导致使用同样长度的码字表示所有的符号会造成浪费。通过熵编码，针对不同的语法元素分配不同长度的码元，可以有效消除视频信息中由于符号概率导致的冗余。

在视频编码算法中常用的熵编码方法有变长编码和算术编码等，具体来说主要有**上下文自适应的变长编码(CAVLC)和上下文自适应的二进制算术编码(CABAC)。**