

视频获取与处理



图 5-1 第 5 章内容结构图

本章学习目标：

- 理解视频的相关基本概念。
- 理解视频的基本参数。
- 了解视频数字化的常见硬件设备。
- 概括视频的数字化过程。

- 分析分辨率、帧速率及视频编码格式对视频数据量的影响。
- 理解音视频封装及常见封装格式。
- 理解视频编码与封装的对应关系。
- 应用主流工具进行视频编辑处理。
- 应用主流工具进行视频效果处理。
- 探讨 AI 技术对视频编辑与处理的影响。

5.1 视频的基础知识

5.1.1 基本概念

第4章讲过,将动态视觉信息放到计算机中,可以存储为动画和视频两种媒体形式。在多媒体技术中,通过实时摄取方式获得的连续动态的图像序列就称为视频(video)。把视频放入计算机中形成数字文件,即数字化视频。

视频的最基本单位是帧(frame)。当连续的图像变化每秒超过24帧时,根据视觉暂留原理,人眼无法辨别单幅的静态图像,看上去是平滑连续的视觉效果。视频整体的播放效果就由序列图像帧的清晰度和流畅度决定。

模拟视频是早期的视频技术,随着数字技术的不断发展,数字视频逐渐取代了模拟视频。数字视频是将模拟视频信号经过数字化处理转换成计算机能够处理的数字信号,从而可以在计算机上进行编辑、存储、传输和播放的视频形式。数字视频具有图像质量稳定、易于编辑和传输、可以长时间存储而不会出现质量损失等优点,已经成为现代多媒体技术的重要组成部分。

5.1.2 基本参数

视频的观看体验是由视频画面的清晰度、流畅度、播放设备及环境等多种因素共同决定的。视频是一种动态视觉媒体,右击图5-2(a)所示的视频文件,可以查看其属性的详细信息,能够查看到图5-2(b)所示的视频信息的基本参数,如时长、帧宽度、帧高度、数据速率和帧速率。

1. 分辨率

分辨率是决定视频质量和观看体验的重要参数,即图像包含的像素数量,它通常以水平像素数和垂直像素数的形式表示,如 1280×720 或 1920×1080 。分辨率越高,视频图像显示的细节就越多,画质也就越清晰。一个视频的分辨率为 1280×720 ,就代表了这个视频的水平方向有1280个像素,垂直方向有720个像素。图5-3列出了常见视频分辨率的对比。

随着技术的发展,视频分辨率的数值越来越大,于是,产生了两种将视频分辨率按照约定俗成的方法进行缩略的命名规则。



图 5-2 视频及基本参数

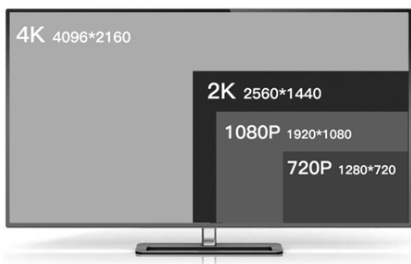


图 5-3 不同视频分辨率的对比

第一种：垂直分辨率及扫描方式的缩略命名。例如 480p, 480I, 720P, 720I, 1080P, 1080I。其中,前面的数字表示垂直像素数,后面的字母是扫描方式的缩写。逐行扫描 (progressive scanning) 缩写为 P, 隔行扫描 (interlaced scanning) 缩写为 I。例如,图 5-3 中的 720P 就是指 1280×720 逐行扫描。达到 720P 以上的分辨率,是高清信号的准入门槛,因此,720P 标准也被称为高清 (high definition, HD) 标准。

传送数字图像时,将每幅图像分解成很多像素,按一个一个像素、一行一行的方式顺序传送或接收,就称为扫描。隔行扫描是将一帧图像分成两场 (从上至下为一场) 进行扫描,第一场先扫描 1、3、5 等奇数行,第二场再扫描 2、4、6 等偶数行,普通的电视机一般都采用隔行扫描。逐行扫描是将各扫描行按照次序扫描,即一行紧跟一行的扫描方式,计算机显示器都采用逐行扫描。随着数字电视技术的发展,隔行扫描方式逐渐被逐行扫描方式取代。

第二种：水平分辨率的缩略命名,水平像素数接近几千,就缩写为几 K。例如 2K (1920×1080)、8K (7680×4320),图 5-3 中的 4K 是由 4096×2160 分辨率缩略而来。

根据视频的像素数量,分辨率可以划分为以下等级。

(1) 标清(standard definition, SD): 标清是视频清晰度的一个基础等级,其物理分辨率一般在 480P 左右。这种清晰度的视频一般应用在一些对画质要求不高的场合,如日常家庭观看、网络视频等。

(2) 高清(high definition, HD): 高清视频的分辨率通常达到 720P 或 1080I,其画面质量相较于标清有了显著提升。高清视频在现代电视、网络视频、蓝光碟等领域得到了广泛应用,能够为用户带来更加清晰、细腻的画面效果。

(3) 全高清(full high definition, FHD): 全高清视频的分辨率一般为 1080P,即 1920×1080 像素。这种清晰度的视频在蓝光碟、高端电视、高端显示器等领域得到了广泛应用,能够为用户带来更加逼真、生动的画面效果。

(4) 超高清(ultra high definition, UHD): 超高清视频的分辨率通常达到 4K 或更高,其像素数量远超全高清。超高清视频在高端电视、电影院、高端显示器等领域得到了广泛应用,能够为用户带来更加细腻、逼真的画面效果,让用户具有身临其境的视觉体验。

随着技术的不断发展,视频清晰度的等级也在不断提升,如 8K 等高分辨率的视频格式为用户带来更加震撼、真实的视觉体验。

2. 画面宽高比

视频画面的宽高比也称为画面比例,是指视频画面的宽度和高度之比。它通常用数字表示,如 4 : 3、16 : 9 等。

不同的宽高比适合不同的场景和设备。例如,4 : 3 是一个比较老的宽高比,常用于 CRT 电视、老式电影、经典游戏等。然而,随着科技的进步和电影制作技术的发展,16 : 9 的宽屏比例逐渐成为主流,适用于大多数高清电视和电影院的银幕,也成为现代电影制作的主要选择。此外,宽高比为 2.35 : 1 甚至更高的超宽屏则创造了更为宽阔的画面空间。对于一些特定的移动设备应用,为了更好地适应竖屏使用习惯,方便用户单手操作和观看,9 : 16 的垂直方向视频比例也非常受欢迎。

画面的宽度 = 水平像素数 \times 像素的长度,画面的高度 = 垂直像素数 \times 像素的宽度。而画面的基本构成单位是像素,像素长宽比是指像素的长度和宽度之比。当图像中的每一个像素不是方形像素时,即像素长宽比不等于 1.0 时,画面的宽高比就不等于水平像素数与垂直像素数之比了,即

$$\text{画面宽高比} = \frac{\text{画面的宽度}}{\text{画面的高度}} = \frac{\text{水平像素数} \times \text{像素的长度}}{\text{垂直像素数} \times \text{像素的宽度}}$$

例如,图 5-4 中的两幅画面宽高比是一样的,但其中,上方画面的像素为正方形像素,而下方画面的像素则为矩形,两幅画面的分辨率不同,观看效果也是不同的。

在 Adobe Premiere Pro(注:本章使用 Adobe Premiere Pro 2023 作为演示工具,以后简称 PR)的时间线上新建序列时,如图 5-5 所示,当像素长宽比选择“方形像素(1.0)”时, 1920×1080 分辨率的画面宽高比为 16 : 9,但遇到一些特殊屏幕的显示需要时,如像素长宽比为“DVCPRO(1.5)”时,画面宽高比 = $1920 \times 1.5 / 1080 \times 1 = 8 : 3$,此时,画面宽

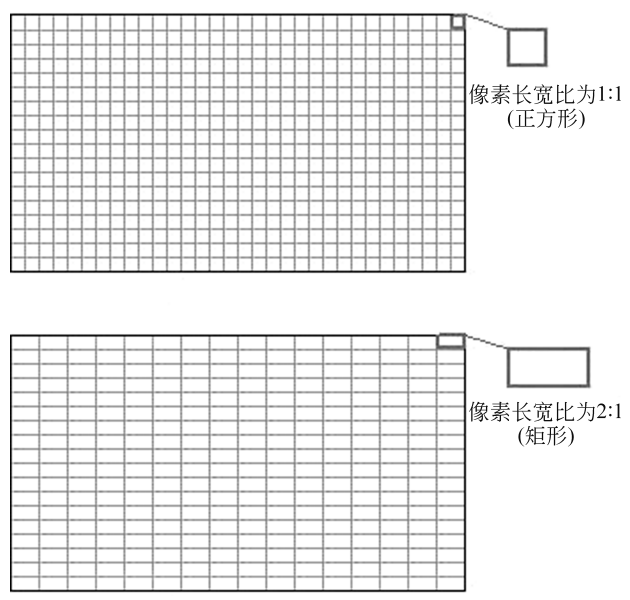


图 5-4 像素长宽比

高比就是 8 : 3 了。因此,即使在帧大小相同的基础上选择不同的像素长宽比,最终的画面宽高比也会不同。

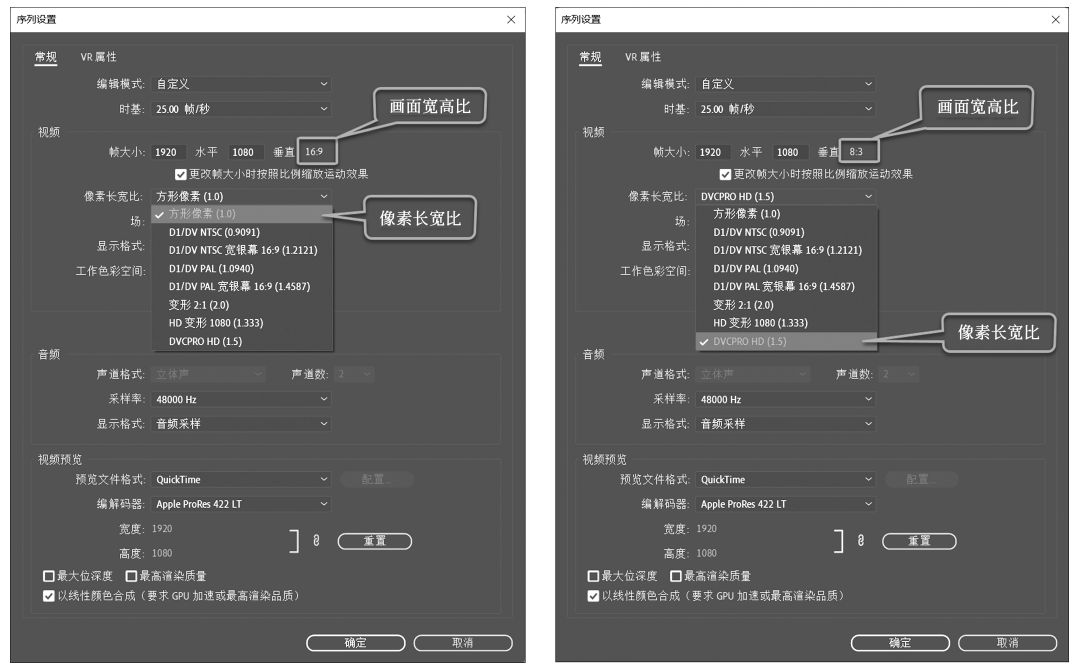


图 5-5 画面宽高比与像素长宽比的设置

如何选择像素长宽比,取决于数字视频将在什么样的终端显示。如果仅仅是在计算机显示器上显示,就选择方形像素,但如果需要在宽屏显示设备上显示,则要选择相应的

像素长宽比,以免发生变形。

3. 帧速率

帧速率表示视频每秒钟显示的图像数量,也称帧频,它是视频流畅度的体现。帧速率通常以 fps(frames per second)表示,如 24fps、30fps、60fps 等,帧速率越高,越可以得到更加流畅、逼真的视觉效果。

PR“序列设置”窗口中的“时基”指的就是帧速率,如图 5-6 所示。进行视频编辑时,可以根据需要调整“时基”来设置合适的帧速率,以获得理想的视频效果。



图 5-6 帧速率的设置

当拍摄的素材帧速率不一致时,视频处理软件会通过插帧或抽帧的方法来匹配序列设置。因此,拍摄和编辑视频时,尽量确保素材的帧速率与编辑序列的帧速率一致。

4. 制式

实现电视的特定方式,称为电视的制式。制式定义了对视频信号的解码方式。

不同制式对色彩的处理方式、屏幕扫描频率等有不同的规定。因此,如果计算机系统处理视频信号的制式与其相连的视频设备的制式不同,会明显降低视频图像的效果,有的甚至根本无法显示图像。

在黑白和彩色模拟电视的发展过程中,分别出现过许多种不同的制式。各国的模拟电视制式不尽相同,制式的区分主要在于帧频、分辨率、信号带宽以及色彩空间的转换关系不同等。模拟电视视频制式标准有以下3种:NTSC制式、PAL制式、SECAM制式。其具体的对比如表5-1所示。由于数字电视是从模拟电视发展而来的,这3种模拟彩色电视制式互不兼容,因此数字电视的格式明显带有各种彩色电视制式的痕迹。而数字电视格式的标准目前并未统一。

表 5-1 3种模拟电视制式的对比

	电 视 制 式		
	NTSC	PAL	SECAM
代表国家	美国、加拿大、日本	中国、德国、英国	法国、俄罗斯
分辨率	720×480	720×576	720×576
画面宽高比	4:3	4:3	4:3
帧频	30	25	25

5.2 视频的数字化

5.2.1 硬件设备

视频的数字化需要借助视频获取硬件设备完成,常见的视频获取硬件设备如图5-7所示。



图 5-7 视频获取硬件设备

首先,可以通过数字摄像机直接产生数字视频信号,存储在介质上,放入计算机中。也可以通过视频采集卡获取数字视频信号。这类采集卡用于将摄像机、录像机、电视机等输出的模拟信号从模拟输入接口输入,将其数字化后再存入计算机。其中,将模拟视频信号数字化并转换为计算机数字视频信号的多媒体卡称为视频捕捉卡或视频采集卡。模拟视频数字化需要经过一系列技术处理:包括颜色空间的转换、扫描方式的转换、分辨率的转换等过程。

5.2.2 数字化过程

视频的数字化过程同样需要经过采样、量化和编码 3 个基本步骤。

1. 采样

采样与量化是将连续的模拟视频信号转换为离散的数字视频信号的关键步骤,为后续的编码、存储和传输奠定了基础。

当进行视频采样时会涉及两个维度:空间采样和时间采样。

空间采样:也称为图像采样,是对图像在空间维度上的离散化。它决定了图像的分辨率,即图像的像素数量。空间采样率越高,图像的像素数越多,视频的分辨率就越高。

时间采样:是对视频在时间维度上的离散化。它决定了视频的帧速率,即每秒钟播放的帧数。时间采样率越高,视频的帧速率越高,动态效果越流畅。

摄像头或图像传感器会以一定的帧速率捕捉场景中的光线,并转换为电信号,这些电信号随后被数字化为像素值,并组成一帧图像,它是构成数字视频的基本单位。

2. 量化

视频中每一帧图像的量化过程与单独的图像量化过程在原理上是相同的。在数字图像处理中,量化是将图像的连续取值(如颜色、亮度等)转换为离散值的过程。对于视频来说,每一帧都可以看作一个静态的图像,因此也需要进行类似的量化处理,可以参考第 3 章图像的量化过程来理解。

具体来说,在视频的每一帧中,像素的颜色和亮度信息通常会被量化为特定的二进制数值,即以一定的位深度来表示每一个像素的信息。位深度决定了每个像素的颜色信息用多少二进制位来表示。24 位深度意味着每个颜色通道(红、绿、蓝)用 8 位二进制表示,共可以表示大约 1670 万种颜色。更高的位深度则意味着更多的颜色级别和更精细的颜色过渡。在广播电视和大多数消费级视频设备中,24 位位深度是常见的选择,因为它在颜色表现和文件大小之间达到了较好的平衡。然而,在电影制作和高端家庭影院等需要更高色彩精度的应用中,可能会采用更高的位深度。

在视频的采样和量化过程中,确定了视频的一些基本参数,包括每帧图像的分辨率(即水平方向和垂直方向的像素数)、帧速率(即每秒传输的帧数)、图像量化时的位深度(每个像素点用多少位二进制数表示颜色信息)以及视频时长。

此时,使用以下公式可以计算出视频采样和量化后的数据量。

视频数据量 = 每帧图像的数据量 × 帧速率 × 视频时长

其中,每帧图像的数据量可以通过以下方式计算

每帧图像的数据量 = 图像分辨率 × 位深度 / 8

通过以上公式计算得到的就是未经压缩的视频数据量。我们试着计算一下:1分钟时长、帧速率为25FPS,分辨率为 1920×1080 ,位深度为24的数字视频,不经压缩的数据量是多少?具体计算公式如下。

示例视频数据量 = 图像分辨率 × 位深度 × 帧速率 × 视频时长 / 8

示例视频数据量 = $1920 \times 1080 \times 24 \times 25 \times 60 / 8 \approx 8\text{GB}$

由此计算可知,采样和量化后的原始视频数据量非常庞大,需要对其进行压缩编码。视频数据中存在大量的冗余数据,包括时间冗余、空间冗余、结构冗余、视觉冗余、知识冗余和数据冗余等,在保证视频质量相同的前提下,挖掘利用的冗余越多,压缩率就越高。

3. 编码

视频的编码格式决定了视频数据如何被压缩、存储和传输。视频编码格式的发展是一个不断演进的过程,随着技术的不断进步和应用场景的不断扩展,新的编码格式不断涌现。

不同的视频编码格式采用了不同的压缩算法和技术,以在压缩效率、图像质量、解码复杂度等方面达到不同的平衡。例如,H.264/AVC被广泛认为是目前最有效率的视频压缩技术之一,适用于高清视频、视频会议、网络流媒体等多种应用(注:H.264和AVC是同一种视频编码标准的两种不同命名,前者是由国际电信联盟电信标准化部门的视频编码专家组制定的标准,而AVC(advanced video coding)是国际标准化组织/国际电工委员会的运动图像专家组命名的同一标准);而H.265/HEVC则提供了更高的压缩效率,适用于高清和超高清视频,特别是在带宽受限的网络环境中(注:H.265/HEVC是H.264/AVC的继任者)。

在PR中导出视频时,可以根据需要选择不同的编码格式。如图5-8所示的“格式”下拉菜单中列出了目前主流的视频、音频和图像的编码格式,其中就包括H.264、HEVC(H.265)、MPEG4、QuickTime、Windows Media等常见的视频编码格式。选择合适的编码格式需要考虑视频内容、目标平台、带宽要求、存储空间等多个因素。

4. 封装

我们平时观看的视频文件经常是一种多媒体文件,它结合了视频流和音频流,以及一些其他可能的数据流,如字幕等。音视频封装就是将已经编码压缩好的视频数据和音频数据按照一定的格式放到一个文件中,以便于播放和传输,也可以理解为将视频和音频数据打包成一个整体的文件,而这个文件需要遵循一定的规范和标准,以确保不同的设备和软件都能够正确地解析和播放,这个过程可以用图5-9简要表示。

视频数据主要负责画面的展示,包括图像的序列和动态效果。而音频数据则负责声音的表现,包括对话、音乐、音效等。这两部分数据在视频文件中是同步的,以确保观看

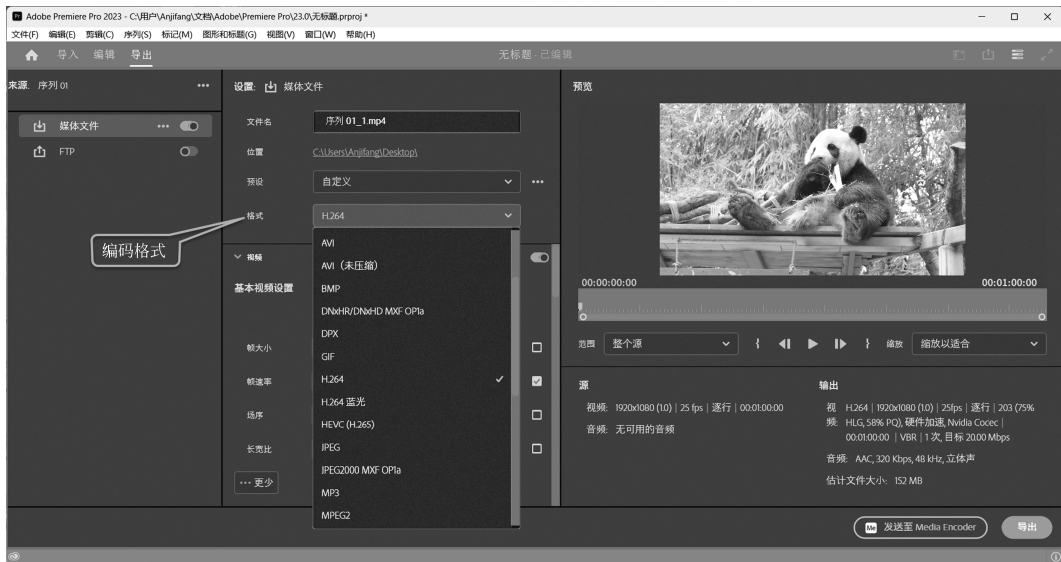


图 5-8 选择编码格式

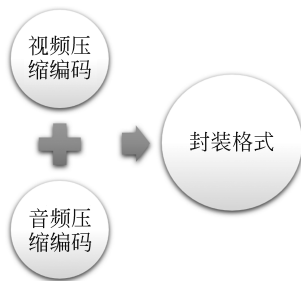


图 5-9 封装的简要表示

视频时画面和声音能够同步播放,提供完整的视听体验。

在封装过程中,视频和音频数据被封装在一个容器内,这个容器可以理解为一个存放视频和音频数据的“盒子”。常见的封装格式(容器格式)包括 MP4、AVI、MKV 等。这些格式规定了如何组织和存储视频、音频以及其他可能的数据流(如字幕等)。

在 PR 中导出视频时,选择不同的视频编码格式通常会导致导出的文件格式不同。例如,如果选择 H.264 编码格式,则通常会导出为 MP4 文件格式,图 5-8 中就是这种情况。这是因为 MP4 是一个广泛支持的容器格式,可以很好地与 H.264 编码的视频配合使用。类似地,如果选择 QuickTime 编码格式,则会导出为 MOV 文件格式。在平时生活中,一些常见的视频编码格式与封装文件格式的对应关系如表 5-2 所示。

导出视频时,最好选择与目标播放平台和观众设备兼容性最佳的视频编码格式和容器格式组合。还可以考虑使用专业的视频编码软件或转换工具来进一步优化视频导出设置,以获得更精细的控制和更好的结果。