HW 多媒体技术基础

PB19151769　马宇骁

目录

0 绪论

0.1 什么是多媒体和多媒体技术？简述多媒体技术与多媒体信息系统的关系?

多媒体就是多重媒体的意思，媒体是人与人之间实现信息交流的中介。因此多媒体可以理解为直接作用于人感官的文字、图形、图像、动画、声音和视频等各种媒体的统称，即多种信息载体的表现形式和传递方式。

多媒体技术，就是利用电脑把文字、图形、影象、动画、声音及视频等媒体信息都数位化，并将其整合在一定的交互式界面上，使电脑具有交互展示不同媒体形态的能力。

多媒体技术与多媒体信息系统关系：

1. 多媒体技术是建立多媒体系统的基础;
2. 多媒体系统的建立与应用又反过来促进多媒体技术的不断完善、发展;
3. 多媒体技术与多媒体系统都会随着与它们相关的其它技术的进步而不断向前发展。

0.2 试归纳叙述多媒体关键特性以及这些特性之间的关系

多媒体的关键特性主要包括信息载体的多样性、交互性和集成性这三个方面，这既是多媒体的主要特征，也是在多媒体研究中必须解决的主要问题。信息载体的多样性是相对于计算机而言的，指的就是信息媒体的多样化，有人称之为信息多维化; 多媒体的第二个关键特性是交互性，多媒体系统将向用户提供交互式使用、加工和控制信息的手段，为应用开辟更加广阔的领域，也为用户提供更加自然的信息存取手段; 多媒体的集成性主要表现在两个方面，一是多媒体信息媒体的集成，二是处理这些媒体的设备与设施的集成。

信息载体的多样性是集成性的基础，没有多种信息媒体，也就无法进行多媒体信息的集成化处理; 而处理多媒体的设备与设施的集成性是实现交互性的前提，没有系统、网络、软硬件设施的集成，就无法为用户交互式使用、加工和控制信息提供平台。

0.3 媒体的结合为什么会产生“感觉相乘”的效果？试举例对此加以说明

多媒体的作用在很大程度上是媒体之间结合产生的影响。这种结合可以是低层次的，如在显示窗口中提供多种媒体信息片断，并将视觉、听觉相互结合，造成一种比较适合的媒体表现环境; 也可以是高层次的，由各种媒体组成完全沉浸的虚拟空间，但应该如何结合现在还缺乏理论上的指导。媒体之间可以相互支持，也可以相互干扰。如果媒体之间是相互支持的关系，则这种媒体结合所产生的效果就是“感觉相乘”效应。

例子：VR 技术在为了更好的沉浸感，考虑到了虚拟的听觉和视觉的感觉到结合。

0.4 试分析人机交互方式的变化趋势

通过传感器直接或间接与人接触获得感知信息；通过建立模型对感知信息进行分析与识别；对分析结果进行推理达到感性的理解；将理解结果通过合理的方式表达出来。也就完成了人机交互的全过程。

人机交互是指通过计算机输入、输出设备，以有效的方式实现人与计算机对话的技术。人机交互技术包括机器通过输出或显示设备给人提供大量有关信息及提示请示等，人通过输入设备给机器输入有关信息，回答问题及提示请示等。人机交互技术是计算机用户界面设计中的重要内容之一。

因此，人机交互方式就用户界面的具体形式而言, 过去经历了批处理、联机终端 (命令接口)、(文本) 菜单等多通道—多媒体用户界面和虚拟现实系统。就用户界面中信息载体类型而言，经历了以文本为主的字符用户界面 (CUI)、以二维图形为主的图形用户界面 (GUI) 和多媒体用户界面，计算机与用户之间的通信带宽不断提高。就计算机输出信息的形式而言, 经历了以符号为主的字符命令语言、以视觉感知为主的图形用户界面、兼顾听觉感知的多媒体用户界面和综合运用多种感观 (包括触觉等) 的虚拟现实系统。

二十一世纪后，以虚拟现实为代表的计算机系统的拟人化, 以及以手持电脑为代表的计算机的微型化, 是当前计算机的重要的发展趋势。以鼠标和键盘为代表的 GUI 技术不再是主导, 而是利用人的多种感觉和动作通道 (如语音、手写、姿势、视线、表情等输入) 的方式与计算机环境进行交互, 将大大提高人机交互的自然性和高效性。

1 多媒体计算机系统

1.1 多媒体计算机系统概述

1.1.1 试从计算机组成（硬件）角度阐述 PC、MPC、图形工作站、GPU 服务器、超级计算机的差异

计算机硬件组成包括：电源，主板，CPU，内部寄存器，硬盘，显卡，输入硬件，输出硬件等。

PC 正常情况下包含上述硬件；MPC 在一般个人计算机的基础上，通过扩充使用视频、音频、图形处理软硬件来实现高质量的图形、立体声和视频处理能力。MPC 联盟规定多媒体计算机包括 5 个基本组成部件：个人计算机（PC）、只读光盘驱动器（CD-ROM）、声卡、Windows 操作系统、音箱或耳机。同时对主机的 CPU 性能，内存（RAM）的容量，外存（硬盘）的容量以及屏幕显示能力也有相应的限定。

图形工作站也是由这几部分硬件组成，但从显卡角度，就市面上的显卡来说，就不是 PC 常用的 AMD 的 RX6800XT，Nvidia 的 RTX3090 等显卡，而是用专业显卡如 tesla 的卡等。CPU 方面也可以不使用 AMD 的 Ryzen9 5900X 或者 intel 的 12900KF 等，而去使用线程撕裂者等 CPU。

GPU 服务器是基于 GPU 的应用于视频编解码、深度学习、科学计算等多种场景的快速、稳定、弹性的计算服务。GPU 服务器主板与 PC 不同，显卡的种类也是专业显卡如 V100，P100。可以没有显示器等输出设备。

超级计算机的硬件组成与个人计算机组成也基本相同，但 CPU 等为专用定制，不具有通用性与量产性。

1.1.2 截止 2022 年 6 月，神威太湖之光和天河二号在 top500 的排名？

神威太湖之光: 6

天河二号: 9

1.1.3 Intel 第 12 代酷睿处理器支持 ×12 PCIe 4.0。请问该总线最高吞吐量是多少？

23.631GB/s

1.1.4 Intel 第 12 代酷睿处理器集成的 Wi-Fi 5 的通信标准是哪个？

IEEE802.11n, 150Mbps

1.1.5 试对比单机系统的视频处理与基于云的视频处理的优缺点

- 从服务器的采购成本来看

云服务器的存放模式需要向云平台供应商（阿里云、腾讯云等）需要采购云服务器。且需要每年都支付费用，不然系统将无法使用。

使用本地系统数据存放本地则需要采购服务器硬件，一般是电脑主机即可，成本较低。如果硬气使用没有问题，后期可以不用产生其他费用，不受服务商约束。但需要提到的一点是，本地服务器需要专人维护，包含硬件的维修、网络的连接。

- 从数据的安全性方面来看

云服务器存放形式：数据全部存储在云端服务器，有专业的云平台供应商进行托管维护。目前所有提供云服务器服务的提供商都有专业的运维团队和安全专家，以阿里云为例：阿里云的安全团队由信息安全、安全审计、物理安全 3 个团队组成，其研究成果取得第三方认证：ISO27001 国际认证、云安全国际认证，并且通过国家安全等级保护测评，为阿里云服务器提供了超高的安全技术保障。再者，提供软件产品的供应商也有责任进行云服务器的维护，保障客户数据的数据安全。同样的，云服务器支持数据实时同步备份之外，也可支持手动备份，备份数据也放在云端，为数据安全提供双重保障。

本地服务器存放数据：数据存放在本地硬盘中，排除人为拷贝、中病毒等情况以外，基本上不存在数据泄露的问题。不需要连接互联网就能访问，别人无法通过远程方式等任何方式对数据进行盗取和复制。但是这种储存方式对服务器的硬件要求很高，如果本地服务器损坏或者数据丢失就无法找回。

- 从后期服务方面来看

这部分主要是针对于依托于云服务器的各类软件产品的维护服务来说的：以咱们诺怀云物管为例，系统一般都是在晚上自动升级/更新/维护，不耽误工作时间段客户的正常使用。由于升级成本低，用户使用的系统可以快速更新。会提供专属售后服务人员，包括线上咨询、电话沟通、远程演示等多渠道方式进行问题处理。从效率上来说的话，线上反馈及时响应，无需赶赴客户现场，直接通过客户反馈排查原因并进行处理，效率更高。

采用本地服务器，基本不涉及到后续服务，也就不产生维护费。当然了，如果产品需要升级或者维护，则是需要收取维护费的（还得看软件产品的服务商还有没有在继续做维护升级）。需派

维护人员去客户现场进行升级更新处理，提交维护申请加上路上的时间，响应时间会进一步的延长。

- 从功能拓展方面来看

采用云服务器的软件产品可更好的支持二次功能开发，因为属于采用的是更为先进的开发技术和开发框架，为后期的二次开发预留了更多的接口。此外，还可对接多种第三方硬/软件系统，包括门禁、监控、停车场、税控系统、财务系统等，应对现在不少物业企业实现智能化转型提供了更为便捷的窗口，为智能社区改造提供更多的可实施方案。同时还能更好的支持开发多种缴费渠道，比如微信缴费、支付缴费、POS 缴费等，拓展线上缴费的渠道。

而本地服务器的话在这一块则局限性比较大，它主要依赖于厂商本身提供的软件功能，对于新功能的新增比较受限。且支持的缴费方式比较有限，集中于线下缴费做记录之用。从软件使用便捷性来讲

依托云服务器的软件产品无需下载安装，访问专属网址即可。换电脑也不用担心设置匹配的问题，而导致的系统不可用。同时，软件系统使用不受地域限制，出差、居家办公均可。在这这类的软件产品能够快速新增项目，对于有多项目、异地协同办公需求的企业更加适合。

而本地服务器依托的单机版软件则需要下载安装包，更换电脑需要重新下载。且只能在局域网内办公。不过这对于单项目管理、不涉及到出差或是领导不喜欢查看报表的企业来讲，影响也比较有限。

1.2 从基于 CPU 的计算到异构计算

1.2.1 简述 CPU 和 GPGPU 之间的联系和区别

CPU：复杂任务，核少，做串行，计算能力只是 CPU 很小的一部分，处理复杂逻辑；GPU：简单任务，核多，做并行（大吞吐量），做显卡的图象单元计算。但以核心为例，CPU 的核心比较重，可以用来处理非常复杂的控制逻辑，预测分支、乱序执行、多级流水等等 CPU 做得非常好，这样对串行程序的优化做得非常好；但是 GPU 的核心就是比较轻，用于优化具有简单控制逻辑的数据并行任务，注重并行程序的吞吐量。CPU 的核心擅长完成多重复杂任务，重在逻辑，重在串行程序；GPU 的核心擅长完成具有简单的控制逻辑的任务，重在计算，重在并行。

它们分别针对了两种不同的应用场景。CPU 需要很强的通用性来处理各种不同的数据类型，同时又要逻辑判断又会引入大量的分支跳转和中断的处理。这些都使得 CPU 的内部结构复杂。而 GPU 面对的则是类型高度统一的、相互无依赖的大规模数据和不需要被打断的纯净的计算环境。GPU 采用了数量众多的计算单元和超长的流水线，但只有非常简单的控制逻辑并省去了 Cache。而 CPU 不仅被 Cache 占据了大量空间，而且还有有复杂的控制逻辑和诸多优化电路.

1.2.2 多媒体计算具有哪些基本特点？

1. 集成性: 能够对信息进行多通道统一获取、存储、组织与合成。
2. 控制性: 多媒体技术是以计算机为中心，综合处理和控制多媒体信息，并按人的要求以多种媒体形式表现出来，同时作用于人的多种感官。

3. 交互性: 交互性是多媒体应用有别于传统信息交流媒体的主要特点之一。传统信息交流媒体只能单向地、被动地传播信息，而多媒体技术则可以实现人对信息的主动选择和控制。

4. 非线性: 多媒体技术的非线性特点将改变人们传统循序性的读写模式。以往人们读写方式大都采用章、节、页的框架，循序渐进地获取知识，而多媒体技术将借助超文本链接 (Hyper Text Link) 的方法，把内容以一种更灵活、更具变化的方式呈现给读者。

5. 实时性: 当用户给出操作命令时，相应的多媒体信息都能够得到实时控制。

6. 信息使用的方便性: 用户可以按照自己的需要、兴趣、任务要求、偏爱和认知特点来使用信息，任取图、文、声等信息表现形式。

7. 信息结构的动态性: “多媒体是一部永远读不完的书”，用户可以按照自己的目的和认知特征重新组织信息，增加、删除或修改节点，重新建立链。

1.2.3 现代 CPU 提供了哪些可加速多媒体计算（如视频压缩编码算法）的技术？

提供算力给交互界面设计打下基础；更好的图像处理、音频信号处理；将多台异地互联的多媒体计算机协同工作，更好的实现信息共享，提高工作效率等

1.2.4 试述适用于 GPU 的算法应有哪些特点？

1. 高度并行

2. 计算密集型

3. 控制简单

4. 多个阶段执行

5. 浮点型运算

1.2.5 在异构处理器上进行软件开发有哪些特点？

异构处理器开发一般需要处理的数据量比较大，数据以数组或矩阵形式有序存储，并且对这些数据要进行的处理方式基本相同。软件包，框架，通信以及其他一些体系机构上的问题，目前存在者多中标准。需要使用统一标准如 OpenCL 实现在不同硬件厂商的平台的兼容的编写程序。

2 媒体处理技术

2.1 数字音频处理技术

2.1.1 音频信号的频率范围大约是多少？话音信号频率范围大约是多少？

音频信号的频率范围：20Hz-20kHz；
话音信号频率范围：300Hz-3000Hz.

2.1.2 简述声音数字化的过程，并解释该过程有哪些重要参数

通过话筒以及相关电压放大电路把声波转换成电压的波形。通过“采样”和“量化”可以实现模拟量的数字化，这个过程称为“模数转换”（A/D 转换），承担转换任务的电路和芯片称为“数模转换器”（ADC）

采样就是按一定的频率，即每个一小段时间，测得模拟信号的模拟量值。

采样时测的的模拟电压值，要进行分级量化。方法是按整个电压变化的最大幅度划分成几个区段，把落在某区段的采样到的样品值归成一类，并给出相应的量化值。

通过采样和量化，一个连续的波形变成了一系列二进制数字表示的数据。数字化的声音的质量取决于采样频率和量化分级的细密程度。量化的分辨率越高，所得数字化的声音的保真程度也越好，数据量也越大。

- 重要参数：

1. 每秒钟需要采集多少个声音样本，也就是采样频率 (fs) 是多少；

2. 每个声音样本的位数 (bit per sample，bps) 应该是多少，也就是样本精度。

2.1.3 简析用 2 个声道模拟出 3D 环绕声音效的原理

这是通过建立起一个能够模仿我们人类听音经验的非常具体的音频录制环境来实现的。通过将一对麦克风之间的距离设置成和人类两耳之间相同的距离，我们就可以复制我们的耳朵所能捕获的声波，并对声音进行处理。事实上，实现双声道虚拟环绕声技术最好方法之一就是使用一种叫做“仿真人头”的设备。仿真人头使用复合橡胶做出人头的形状，并分别在两个耳朵的位置安装两个小型的麦克风。当我们需要考虑现实生活中声波在到达耳膜之前需要经过人体某些部位的传导和反弹时，颈部和躯干的加入就非常必要了。有些人会选用另一种更为简单的实现方法：将一对小型麦克风像耳塞一样贴在一个真人的耳朵内部。还有些设备只拥有一对模拟的耳朵和配对的麦克风。

A3D Surround 吸收了 A3D 技术和环绕声解码技术 (如 Dolby 的 ProLogic 和 AC-3)，创建一个围绕听者的 5 组音频流的声场，即产生五个“虚拟音箱”，它实际上是经过 A3D Surround 处理后用两个音箱播放出来的。利用仿声学原理，根据人耳对各空间方向声音信号函数的反应不同，对双声道立体声信号中的反射声、回声等信号提取出后进行技术处理。尽管这些信号仍来自前方，但给人的错觉是来自四面八方。其过人之处是只使用两只普通音箱，在无须杜比编码前提下，可产生出仿 3D 环绕声五声道的放音效果，有如音乐厅的身临其境感觉。当前在电脑多媒体“家庭影院”系统中使用的 Vivid 3D Pro 就是 SRS 技术的典型应用。

2.1.4 视频会议软件中，人说话的声音可以清晰传递，但是背景音乐往往存在较大失真。试分析可能的原因

在视频会议软件的采集声音算法设计时应该是以尽可能使得人的声音清晰为目的设计的，因此，由于话音的信号频率范围和背景音乐对应的频率范围不相同，但采样频率是以话音为标准，背景音乐就会有较大失真。

2.1.5 选择采样频率为 22.050 kHz 和样本精度为 16 比特的录音参数。在不采用压缩技术的情况下，计算录制 2 分钟的立体声需要多少 MB 存储空间 (1MB=1024×1024B)

数据传输率 = 采样频率 * 量化位数 * 声道数/8 = 22.05*16*2/8 = 88.2KB/s
故，两分钟：
数据量 = 2*60*88.2 = 10584KB 大约为 10.34MB

2.1.6 试分析音频编解码器芯片中的滤波器设计指标和人听觉感知特性的关系

音频编解码器芯片中的滤波器设计应该是尽量去调节不同频段使得声音效果被适当的调整，使得人耳的听觉感受更丰满。避免在不同频段过低或者过高。

频段	人耳的听觉感受		
	过低	丰满	过高
6-20kHz	韵味失落	色彩鲜明富于表现力	尖噪、嘶哑、刺耳
600Hz-6kHz	暗淡、朦胧	明亮、清晰	呆板
200-600HZ	空虚无力	圆润有力	生硬
20-200HZ	苍白单薄	丰满、混厚、深沉	浑浊不清渗

2.1.7 简要比较波形音频与 MIDI

波形音频，传输的是声音信号；MIDI 只是控制音高、节奏与响度的指令。

一条条高高低低的线条就是音符，高度代表音符的音高，而长度代表音符的时值。竖线表示音符的强弱。但是只有 MIDI 文件（乐谱），还不能发出声音。这个时候我们需要在 MIDI 文件上外挂一个“音源”，相当于是给乐谱配上了乐器。最后，音源（乐器）就可以依照 MIDI 指令（乐谱）演奏，而生成出来的声音就被混缩成一个直接传递声音信号的波形文件——音频。

2.2 数字图像/视频处理技术

2.2.1 一幅 256 色的图像的颜色深度是多少？

$256 = 2^8$ ，深度是 8.

2.2.2 什么是真彩色和伪彩色？伪彩色有哪些应用？

- 真彩色

真彩色是指图像中的每个像素值都分成 R、G、B 三个基色分量，每个基色分量直接决定其基色的强度。

- 伪彩色

伪彩色图像的含义是，每个像素的颜色不是由每个基色分量的数值直接决定，而是把像素值当作彩色查找表 (color look-up table，CLUT) 的表项入口地址，去查找一个显示图像时使用的 R，G，B 强度值，用查找出的 R，G，B 强度值产生的彩色称为伪彩色。

彩色查找表 CLUT 是一个事先做好的表，表项入口地址也称为索引号。例如 16 种颜色的查找表，0 号索引对应黑色，... ，15 号索引对应白色。彩色图像本身的像素数值和彩色查找表的索引号有一个变换关系，这个关系可以使用 Windows 95/98 定义的变换关系，也可以使用你自己定义的变换关系。使用查找得到的数值显示的彩色是真的，但不是图像本身真正的颜色，它没有完全反映原图的彩色。

2.2.3 简要说明 CRT、等离子电视、LCD、LED、OLED 发光的原理

- CRT

结构是一根真空管，里面有一个或多个电子枪，电子枪射出电子束，电子束射到真空管前屏幕表面的内侧时，屏幕内侧的发光涂料受到电子束的击打而发光产生图像的技术。由于荧光粉被点亮后很快会熄灭，所以电子枪必须循环地不断激发这些点。

- 等离子电视

利用气体放电的显示技术，其工作原理与日光灯很相似。它采用了等离子管作为发光元件，屏幕上每一个等离子管对应一个像素，屏幕以玻璃作为基板，基板间隔一定距离，四周经气密性封接形成一个个放电空间。放电空间内充入氖、氙等混合惰性气体作为工作媒质。在两块玻璃基板的内侧面上涂有金属氧化物导电薄膜作激励电极。当向电极上加入电压，放电空间内的混合气体便发生等离子体放电现象。气体等离子体放电产生紫外线，紫外线激发荧光屏，荧光屏发射出可见光，显现出图像。当使用涂有三原色（也称三基色）荧光粉的荧光屏时，紫外线激发荧光屏，荧光屏发出的光则呈红、绿、蓝三原色。当每一原色单元实现 256 级灰度后再进行混色，便实现彩色显示。等离子体显示器技术按其工作方式可分为电极与气体直接接触的直流型 PDP 和电极上覆盖介质层的交流型 PDP 两大类。目前研究开发的彩色 PDP 的类型主要有三种：单基板式（又称表面放电式）交流 PDP、双式（又称对向放电式）交流 PDP 和脉冲存储直流 PDP。

- LCD

LCD 屏幕需要背光源（back-light），背光源为白光，白光穿透有颜色的薄膜就能显示出彩色。但屏幕显示内容会变化，也就是需要调节背光，这通过液晶层（liquid-crystal）来实现。而在这之前，我们先给屏幕加两张偏光片（光的偏振呢，就只能自己去翻物理书了，我也不懂），一张竖直、一张水平。背光通过第一张偏光片时，只有与偏光片偏振方向相应的部分光能通过，但由于方向与第二张偏光片垂直，所以无法通过，屏幕是黑的。然后就用液晶来调整第一张偏光片后的偏正光方向，使其能够通过第二张偏光片。

液晶，即液态晶体（Liquid Crystal，LC），它具有独特的光学性质，光线会沿着液晶的晶体方向传播，所以液晶会扭曲偏振光的方向。通过合适的排列，就能旋转通过第一张偏光片后的偏

振光，使其通过第二张偏光片。液晶分子排列方向被静电场改变，那通过的光线旋转方向自然也发生改变。所以，还需要在液晶层两侧加一对电极板，并通过调整电极板之间的电压来控制液晶分子排列的扭曲程度，从而控制通过第二块偏光片的光量，最终呈现出各种各样的颜色。

- LED

发光二极管的核心部分是由 p 型半导体和 n 型半导体组成的晶片，在 p 型半导体和 n 型半导体之间有一个过渡层，称为 p-n 结。如下图，它主要由支架、银胶、晶片、金线、环氧树脂五个部分所组成。当电流通过晶片时，N 型半导体内的电子与 P 型半导体内的空穴在发光层剧烈地碰撞复合产生光子，以光子的形式发出能量（即大家看见的光）。

- OLED

有机发光二极管（OrganicLight-Emitting Diode，OLED），又称为有机电激光显示、有机发光半导体（OrganicElectroluminesence Display，OLED），是指有机半导体材料和发光材料在电场驱动下，通过载流子注入和复合导致发光的现象。OLED 在电场的作用下，阳极产生的空穴和阴极产生的电子就会发生移动，分别向空穴传输层和电子传输层注入，迁移到发光层。当二者在发光层相遇时，产生能量激子，从而激发发光分子最终产生可见光。

2.2.4 为什么模拟黑白电视和模拟彩色电视的信号带宽均在 6MHz 左右？

对于 640×480 的格式在 30 万像素的情况，其水平清晰度为 320 线，垂直分解力为 240 线。通常厂商取水平分解能力作为线数。420 线彩色 CCD 摄像头，380 线黑白 CCD 摄像头。

N = 420 或 380

- PAL 制、SECAM 制

fs = 625 * 25 * N ≈ 6MHz

- NTSC 制

fs = 525 * 29.97 * N ≈ 6MHz

2.2.5 以 CIE-1931 为例，简述对色域（Color Gamut）的理解

CIE 使用的三原色波长分别为 700nm (R)、546.1nm (G)、435.8nm (B) (因为这三种波长能够精确的产生)，对应的光谱三刺激值如图1所示。

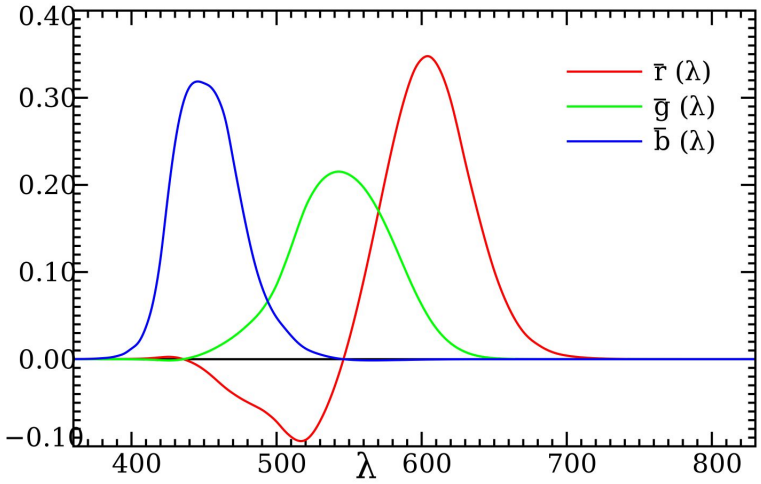

图 1: 光谱三刺激值

光谱三刺激值与色品坐标的对应关系如下：

$$r = \frac{\bar{r}}{\bar{r}+\bar{g}+\bar{b}}, g = \frac{\bar{g}}{\bar{r}+\bar{g}+\bar{b}}, b = \frac{\bar{b}}{\bar{r}+\bar{g}+\bar{b}}$$

从光谱三刺激值计算出其对应的色品坐标，最后连接各个波长的色品坐标形成光谱轨迹如图2。

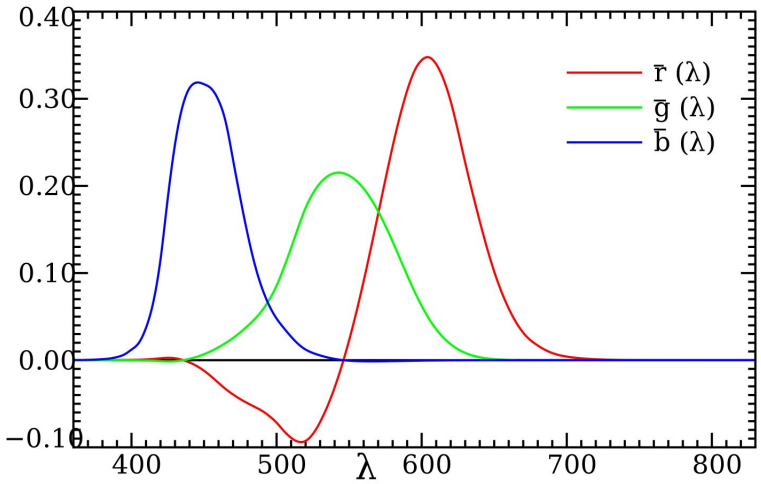

图 2: CIE 1931

由此，可以理解色域：色域是对一种颜色进行编码的方法，也指一个技术系统能够产生的颜色的总和。在计算机图形处理中，色域是颜色的某个完全的子集。颜色子集最常见的应用是用来精确地代表一种给定的情况。例如一个给定的色彩空间或是某个输出装置的呈色范围。

2.2.6 从色域的角度解释为何 HDR10 会流行

HDR10 主要的规格包括了以下几样：

1. EOTF (Electro-Optical Transfer Function)：SMPTE ST 2084
2. Color Sub-sampling：4:2:0（压缩影片）
3. Bit Depth：10 bit
4. 色域：ITU-R BT.2020
5. Metadata（元数据）：SMPTE ST 2086、MaxFALL、MaxCLL

其中 10 bit 的影像比起以往 8 bit 影像可以提供更多光暗层次的数据；BT.2020 也都比全高清时期的 BT.709 色域更广阔，可以显示更多色彩。HDR10 采用的 BT.2020 色域比起以往 HDTV 时期的 BT.709 色域广阔很多，HDTV 的 8 bit 色深可以显示 1 千 6 百万种色彩，如果到了 10 bit 的话就可以显示高达 10 亿种颜色。

2.2.7 简述相加混色模型和相减混色模型的区别和联系

- 相加混色模型

相加混色，就是我们常说的 RGB 模式，适用于显示器等发光体的颜色显示。以黑色介质为基础，通过光源三原色不同比例的亮度叠加，来获得各种不同的颜色。

- 相减混色模型

即 CMY 色彩模式是基于固有色光吸收/反射原理定义，适用于各种印刷媒介，以白色介质为基础，通过印刷三原色（青、品红、黄）不同比例的油墨混合，吸收原始色光中的对应波长，从而获得各种不同的颜色的反射效果，也就是所谓的相减色。例如：黄色颜料是从入射白光中吸收蓝光而反射红光和绿光所致。

- 联系

相加混色模型一般运用于计算机制作出的成果想要打印的时候就需要进行不同色彩之间的转换成相减混合模型才能印刷出人眼看到的样子。

2.2.8 D-Sub 接口某根线接触不良，屏幕泛品红色，请依据色彩空间的知识推测是哪根信号线接触不良

品红色是介于红色和蓝色之间的颜色。故，应该是绿色（GREEN）信号线接触不良。

2.3 数码相机技术

2.3.1 分析 2000 年左右数码相机分辨率达到 200 万像素后迅速取代传统相机的原因

200 万像素已经基本上达到 1080p 的分辨率从分辨率角度已经能与传统相机分不出太大差距；数码相机焦距与变焦，景深/光圈，感光度，测光、曝光与曝光补偿，白平衡，自动对焦等功能集成使得使用者能很方便的拍照调试，降低学习成本，因此也促进了迅速取代传统相机。

2.3.2 试对比智能手机上所用的 RGBY、RYYB、QBC 传感器和单反数码相机中的彩色合成原理

- RGBW 传感器

红蓝绿三色滤色，每一个都会过滤掉部分光线，这就降低了亮度，因为光线损失的比较多。而加入一个白色滤色片，白色是不会过滤光线，投过滤色片照射到 CMOS 上的光线就比较多，这就增加了亮度。而亮度增加，低光环境下的噪点就减少了，这就是华为宣称的提升照片 32% 的亮度（高对比度），低光环境下降低 78% 的彩色噪点。

代价：采集色彩信息的 CMOS 面积下降了 1/4

- RYYB 传感器

将传统 RGBG 拜耳滤镜换成了“RYYB”滤镜，将 2 个绿色像素（G）用黄色像素（Y）替代。代表机型为：华为 P30、P40、Mate 40 Pro+ 等。和 RGGB 相比，RYYB 可以减轻前者在滤色过程中所带来的光之强度折损，可以让进光量提升高达 40%。以华为 P30 Pro 为例，该产品的 ISO 高达 409600，是 iPhone Xs Max 的 64 倍！从而只需一丝亮光就能记录下纯黑环境下的颜色细节。

RYYB 提升了进光量，但变相增加了红色的进光量，从而提升了弱光环境下的表现。由于黄色像素较多，偏色问题将难以避免，同时绿色像素的缺失也会影响饱和度。需要一套更加强大的硬件 ISP 和更为成熟的成像算法支持。

- QBC

是 quad bayer coding 的缩写。这种 sensor 的设计是：每个像素是有四个子像素组成，他们公用一个 color filter。在非 HDR 模式下，四个像素合并成一个输出值，与当年 Nokia 用的 4100 万像素出 8 百万像素的照片的所谓超采样是一个原理，具有降噪，增加动态范围的好处。在 HDR 模式时，会把四个像素分成两组，对角线方向的分到同一组，135 度的那组曝光要短于 45 度的那组，然后再 scale+combine，这样就生成完全 pixel 位置的 HDR 图像。

- 单反数码相机

Sensor3 彩色数字相机需要 3 个单色 sensor 获得彩色图像的 R，G，B 分量，成本较高。单 CCD 获得彩色图像的方法是在 CCD 表面覆盖 1 个只含红、绿、蓝 3 色的马赛克滤镜，对其输出信号通过一定的处理算法实现。这个设计理念 1976 年最初由拜尔提出，所以这种滤镜也被称作拜尔模板 (bayer pattern)。

由于每个像素只感知一种颜色（一种色谱），一个彩色图像感知器的整体感知度比单色（全色）感知器的感知度要低，实际中一般要低到 3 倍。所以，单色感知器在光线不佳的情况下表现要好得多。

2.3.3 请从滤波器角度分析镜头参数 MTF(Modulation Transfer Function) 的意义

MTF 图表描绘镜头从中心到边缘的反差对比度。

- X 轴（水平轴）显示从影像中心到边缘的距离。因此“±0”代表镜头中心，不同的数值代表着朝向镜头边缘方向的距离，以毫米为单位。
- 在 Y 轴（垂直轴）上，最大值是 1，表示光线百分之百完全通过镜头。

MTF 结合了分辨率和对比度两种指标，表示了成像系统将分辨率转换为对比度的能力。随着图像分辨率的升高，对比度会下降，（1mm 的距离内线对更多，因此成像系统造成的模糊带来的影响也更严重，对比度会更差），MTF 表示了成像系统在分辨率升高的情况下保证对比度的能力。

2.3.4 请从人的视觉认知特性角度解释拜尔模板 (bayer pattern) 的设计

Bayer pattern 说的是 COLOR FILTER 的结构, 分为两种：STD Bayer pattern 与 Pair pattern，其中 STD Bayer pattern 的结构是 BG/GR 的，而 Pair Pattern 顾名思义是指 BGBG/GRGR 的结构，即以四行为一个单位，前两行是 BG 的结构，后两行是 GR 的结构 Bayer 格式是相机内部的原始图片, 一般后缀名为.raw。很多软件都可以查看, 比如 PS。我们相机拍照下来存储在存储卡上的.jpeg 或其它格式的图片, 都是从.raw 格式转化过来的。bayer 色彩滤波阵列，由一半的 G，1/4 的 R，1/4 的 B 组成。

根据人眼对彩色的响应带宽不高的大面积着色特点，每个像素没有必要同时输出 3 种颜色。因此，数据采样时，

奇数扫描行的第 1，2，3，4，…象素分别采样和输出 R，G，R，G，…数据；

偶数扫描行的第 1，2，3，4，…象素分别采样和输出 G，B，G，B，…数据。

在实际处理时，每个象素的 R，G，B 信号由象素本身输出的某一种颜色信号和相邻象素输出的其他颜色信号构成。这种采样方式在基本不降低图像质量的同时，可以将采样频率降低 60% 以上。

Bayer RGB 与 RGB Raw 的主要区别在于两者输出前经过的处理不同，Bayer RGB 从 ADC 输出，只经过了 LENS SHADING，GAMMA 等模块处理而后就直接输出，而 RGB Raw 则经过了整个 ISP 模块的处理，最终是经过 YUV422 的数据转化而来的

2.3.5 手机相机可以获取和高端全画幅 DSLR 一样的分辨率，但为何 DSLR 价格昂贵？

在光线等条件都良好的情况下，顺光拍摄一些小场景手机和单反不会造成太大的差别，手机和单反最大的差距就是在景深的控制上。单反镜头的光圈可调节，虽然手机也可以通过软件模拟出光圈可调节效果，通过多摄像头带来可变焦距和人像拍摄模式等，但是在光学结构上依旧存在着差异，因此在景深效果上也是不如单反拍照的；

目前手机已经可以搭配多个不同焦段的摄像头，通过光学和数码的混合变焦方式，实现了模拟单反相机的光学变焦效果，不过在实现的原理上与一般的单反相机借助镜头实现的变焦不同，手机的混合变焦只能配合后期软件算法实现长焦虚化的效果；

单反在画质、色彩解析力、宽容度上相比于手机都是有绝对的优势。而以上的优势是基于单反在影像领域堆料的结果，因此成本造价很高。

2.3.6 请解释为何手机上常使用对比度检测（反差式）对焦？

因为其成本相对较低，且相比相位对焦它对光线要求不高，相比激光对焦可以不增加额外的传感器的成本。对焦的过程就是通过移动镜片来使对焦区域的图像达到最清晰的过程，所以对焦成功以后，直观的感受就是焦点的清晰度最高，而焦点以外的区域表现为相对模糊的状态。采用反差对焦的相机，当我们对准被摄物体时，镜头模组内的马达便会驱动镜片从底部向顶部移动，在这个过程中，像素传感器将会对整个场景范围进行纵深方向上的全面检测，并持续记录对比度等反差数值。找出反差最大位置后，运动到顶部的镜片则会重新回到该位置，完成最终的对焦。所以使用反差对焦的手机在拍照过程中，如果取景框中的物体位置、内容发生了变动，我们的肉眼就可以观察到屏幕上的内容由模糊到清晰再到模糊的过程，有一种镜片在前后推拉的直观感受。因为反差对焦的工作方式是进行对比度检测，因此相机镜片必须要前后移动完整记录所有的图像信息，然后计算对比度最高的位置，才能最终完成对焦。所以反差对焦的一个主要缺点就是耗费的时间较长。反差对焦类似手动调焦的过程：模糊-清晰-模糊，然后重回到清晰的焦距。

2.3.7 解释智能手机相机 OIS 的原理

通过镜头内的陀螺仪侦测到微小的移动，然后将信号传至微处理器，处理器立即计算需要补偿的位移量，然后通过补偿镜片组，根据镜头的抖动方向及位移量加以补偿；从而有效地克服因相机的振动产生的影像模糊。这种防抖技术对镜头设计制造要求比较高，而且成本也相对高一些。光学防抖功能的效果是相当明显的，一般情况下，开启该功能可以降低 2 — 3 档快门速度，使手持拍摄不会产生模糊不清的现象，对于初学者来说效果非常明显。特别在大变焦相机，效果就更为明显了；因为一般变焦越大的情况下，就算是极轻微的抖动也是非常易见的，对于长焦情况下对防抖的功能需求就更大了。

2.3.8 名词解释

Ultra-wide camera 把焦距在 24mm 以下的镜头叫做超广角镜头。

Telephoto camera 长焦镜就是让你可以看得很远的镜头，一般 200mm 以上的镜头便是长焦镜。

3 多媒体数据压缩

3.1 无损数据压缩

3.1.1 名词解释：UTF-8、UTF-16、UTF-32

UTF-8

使用变长字节表示字符的编号

1. 对于单字节的符号，字节的第一位设为 0，后面的 7 位为这个符号的 Unicode 码，因此对于英文字母，UTF-8 编码和 ASCII 码是相同的。

2. 对于 n 字节的符号（n>1）, 第一个字节的前 n 位都设为 1，第 n+1 位设为 0，后面字节的前两位一律设为 10，剩下的位用来存储字符的 Unicode 码。

UTF-16
使用变长字节表示字符的编号

1. 对于编号在 U+0000 到 U+FFFF 的字符，直接用两个字节表示。

2. 编号在 U+10000 到 U+10FFFF 之间的字符，用四个字节表示。

UTF-32
用 4 个字节存储字符的编号.
还需要考虑计算机的端模式（大端、小端）.

3.1.2 某符号的 Unicode 数字编号为 0x4E2D，写出 UTF-8 编号后的 16 进制结果

0x4E2D：0100 111000 101101
$\Longrightarrow$ UTF-8: 1110 0100 10 111000 10 101101
即：0xE4B8AD

3.1.3 已知信源 X: x1, x2, x3, x4, x5, x6，各信源符号的概率依次为 P(X): 0.2, 0.19, 0.18, 0.17, 0.15, 0.1, 0.01。求霍夫曼编码，并计算编码效率

（我怀疑是题写错了，应该有 7 个 x，因为给了 7 个概率）

$$
\begin{aligned}
&H(X) = -\sum_{i=1}^{7} p(x_i) \log_2 p(x_i) \\
&= -(0.2 \times \log_2 0.2 + 0.19 \times \log_2 0.19 + 0.18 \times \log_2 0.18 + 0.17 \times \log_2 0.17 \\
&+0.15 \times \log_2 0.15 + 0.1 \times \log_2 0.1 + 0.01 \times \log_2 0.01) \\
&= 2.609
\end{aligned}
$$

哈夫曼编码为：
x1: 10
x2: 11
x3: 000
x4: 001
x5: 010
x6: 0110
x7: 0111
平均码长：2.72 编码效率：$\frac{2.609}{2.72} \times 100\% = 95.91\%$.

3.1.4 对一个具有符号集 B= b1, b2 = 0, 1 , 设信源产生 2 个符号的概率分别为 P(b1)=0.2, P(b2)=0.8。对二进制数 1001 进行算术编码（结果用十进制数表示）

两个子间隔：[0, 0.2) 和 [0.2, 1]
1001: 0.2 + 0.8*0.2 + 0.8*0.2*0.2 + 0.8*0.2*0.2*0.8 = 0.4176

3.1.5 对信息 0000203300111000006001101111 进行行程（游程）编码

40 12 10 23 20 31 40 16 20 21 10 41

3.2 音频数据的压缩标准

3.2.1 话音编译码器通常可以分成哪三种类型，并举例简述其基本原理

- 波形编译码器

不利用生成话音信号的知识产生而是产生一种重构信号，重构信号的波形和原始话音波形尽可能一致，这种编译码器的复杂程度低。

波形编码代表：PCM（脉冲编码调制）

- 音源编译码器

通过话音波形的信号中提取生成话音的参数，使用这些参数通过话音生成模型重构出话音。在模型中声道被等效成一个随时间变化的滤波器，叫时变滤波器，激励函数是由白噪声，无声话音段激励或者由有声话音段激励。传送的是解码器的信息就是滤波器的规格、发声或不发声的标志和有声话音的音节周期，每 10 20ms 更换一次。

数据率 2.4bps，产生的语音质量很低，可以听懂而已。增加数据率对于话音质量没有用，因为这是由模型限制的，但保密性好。

- 混合编译码器

混合编译码的想法是企图填补波形编译码和音源编译码之间的间隔。波形编译码器虽然可提供高话音的质量，但数据率低于 16 kb/s 的情况下，在技术上还没有解决音质的问题；声码器的数据率虽然可降到 2.4 kb/s 甚至更低，但它的音质根本不能与自然话音相提并论。

为了得到音质高而数据率又低的编译码器，历史上出现过很多形式的混合编译码器，但最成功并且普遍使用的编译码器是时域合成-分析 (analysis-by-synthesis，AbS) 编译码器。

3.2.2 名词解释：听阈，痛阈，频域掩蔽，时域掩蔽

听阈刚能引起人耳听觉反应的最小声音刺激量，称为听阈。

痛阈将各频率的听阈以线段连接，形成听阈曲线。若继续增加声音刺激强度刚能引起人耳不适或疼痛的最小刺激量，称为痛阈。

频域掩蔽同时发出的频率接近的两个纯音，声强低的纯音会被声强高的纯音淹没

时域掩蔽在时间上相邻的声音之间也有掩蔽现象，称为时域掩蔽。产生的主要原因是人的大脑处理信息需要花费一定的时间。

3.2.3 MPEG-1 的层 1、层 2、层 3 编码分别使用了听觉系统的什么特性？

- 层 1

频带相等的子带，使用频域掩蔽特性，每个子带用 6bit 量化。

- 层 2

频带相等的子带，除了使用频域掩蔽特性之外还利用了时间掩蔽特性，低频段的子带用 4 比特，中频段的子带用 3 比特，高频段的子带用 2 比特。

- 层 3

层 3 使用比较好的临界频带滤波器，把声音频带分成非等带宽的子带，除了使用频域掩蔽特性和时间掩蔽特性之外，还考虑了立体声数据的冗余，并且使用了霍夫曼 (Huffman) 编码器。还使用了 MDCT (modified discrete cosine transform) 把子带的输出在频域里进一步细分以达到更高的频域分辨率。

3.3 图像数据的压缩标准

3.3.1 基于 DCT 变换的 JPEG 压缩编码算法的主要计算步骤有哪些？请给出编码框图，并说明此过程中哪些是有损的，哪些是无损的

步骤：

1. 正向离散余弦变换 (FDCT)
2. 量化 (quantization)
3. Z 字形编码 (zigzag scan)
4. 使用 DPCM 对直流系数 (DC) 进行编码
5. 使用 RLE 对交流系数 (AC) 进行编码
6. 熵编码 (entropy coding)

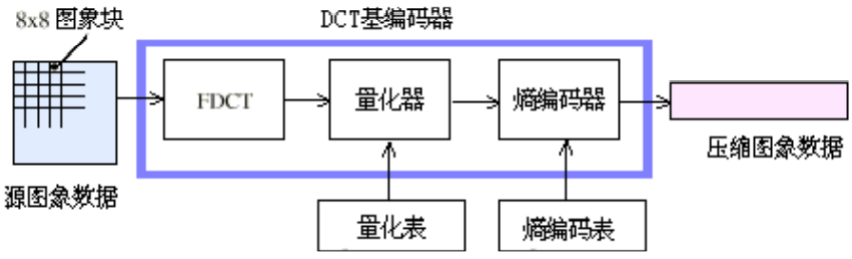

图 3: 编码框图

但做离散余弦变换也会这一些高频成分予以抛弃，从而降低需要传输的数据量，是有损的；
量化是有损的，量化是图像质量下降的最主要原因；
ZIGZAG 无损；
三步编码都有消除冗余的作用，也是有损的。

3.3.2 JPEG 利用了图像的哪几类冗余？

第一种是编码冗余度。例如，DCT 变换，哈夫曼编码，是消除编码冗余度。

第二种是象素间冗余度。例如，直流系数用差分编码就是消除相邻子图象间的灰度 (或亮度) 冗余度。

第三种是心理视觉冗余度。例如，用量化过程，就是利用人眼对各种空间频率，包括亮度、色度、纵、横方向的高频、低频的敏感程度不同，从而降低和消除一部分数据，达到数据压缩的目的，或降低传输位率，同时又不损害心理视觉对图象的主观评价。这就是充分利用心理视觉冗余度对图象数据进行压缩。

3.3.3 为什么 JPEG 使用 DCT 而不是 DFT？

DCT 相比 DFT 有更好的能量聚集效果。

在扩展时，DCT 采用的是对称的形式，DFT 中扩展序列时，是直接采用平移方式引入了不连续的区间，而 DCT 采用对称形式，消除了这一人为的不连续，而通常高频信号隐藏在这人造的不连续中。

而且 DFT 有两种冗余：

- 在有限长度实序列的 DFT 结果中有一半的数据是冗余的；
- 另外由于复数需要表示实部虚部，那么一个复数需要占用两个数据空间。

DCT 的变换结果只有实数部分，没有复数部分。

——————————subsection——————————

3.4 视频数据的压缩标准

3.4.1 什么是视频编码中的运动补偿技术？

运动补偿是通过先前的局部图像来预测、补偿当前的局部图像。

运动补偿是一种描述相邻帧（相邻在这里表示在编码关系上相邻，在播放顺序上两帧未必相邻）差别的方法，具体来说是描述前面一帧的每个小块怎样移动到当前帧中的某个位置去。这种方法经常被视频压缩/视频编解码器用来减少视频序列中的空域冗余。它也可以用来进行去交织（deinterlacing）与以及运动插值（motion interpolation）的操作。

3.4.2 请画出 H.264 的编码框图

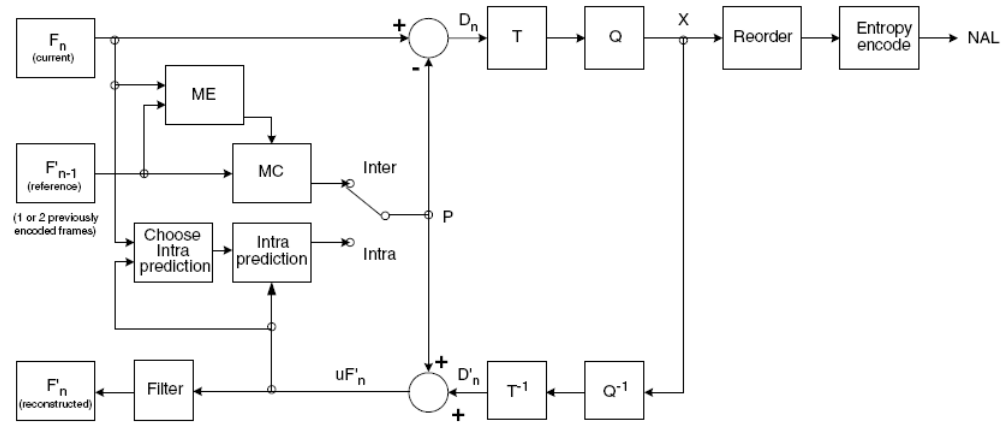

Figure 6.1 H.264 Encoder

图 4: 编码框图

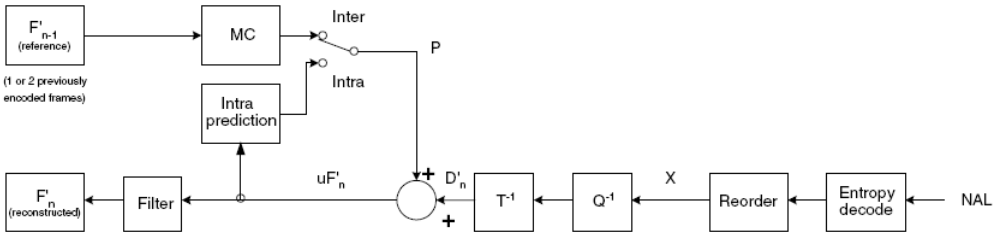

Figure 6.2 H.264 Decoder

图 5: 解码框图

3.4.3 简述 Golomb-Rice 编码的基本原理，并分析在什么情形下该编码能实现高的压缩比

Golomb 编码是一种基于游程编码（run-length encoding,RLE）的无损的数据编码方式。

Golomb 编码是一种分组编码，需要一个正整数参数 m，然后以 m 为单位对待编码的数字进行分组。对于任一待编码的非负正整数 N，Golomb 编码将其分为两个部分：所在组的编号 GroupID 以及分组后余下的部分，GroupID 实际是待编码数字 N 和参数 m 的商，余下的部分则是其商的余数，具体计算如下：

$$q = N/mr = N\%mq = N/mr = N\%m$$

对于得到的组号 q 使用一元编码 (Unary code)，余下部分 r 则使用固定长度的二进制编码 (binary encoding)。

设待编码的非负整数为 N，Golomb-Rice 编码流程如下：

1. 初始化正整数参数 m

2. 取得组号 q 以及余下部分 r，计算公式为：q=N/m,r=N%mq=N/m,r=N%m
3. 使用一元编码的方式编码 q
4. 使用二进制的方式编码 r，r 所使用位数（Golomb-Rice 编码要求参数 m 是 2 的次幂）取 r 的二进制表示的低 log2(m)log2(m) 位，作为 r 的码字。

- 高的压缩比的情况：

当待压缩的数据符合几何分布（Geometric Distribution）时，Golomb 编码取得最优效果。

4 多媒体数据的数字存储

4.1 关于 CD 的黄皮书和红皮书存在哪些重要区别？

- 红皮书

1982 年，Philips 和 Sony 把用“1”和“0”表示的声音信号记录到以塑料为基片的金属圆盘上。由于这种塑料金属圆盘很小巧，所以用了英文 Compact Disc 来命名。这种盘又称为数字激光唱盘 (Compact Disc-Digital Audio，CD-DA) 盘。

- 黄皮书

CD-ROM 存放数字化的文、图、声、象等,1985。但从 CD-DA 过渡到 CD-ROM 有两个重要问题需要解决：1. 计算机如何寻找盘上的数据。2. 要求它的错误率 (10^{-12}) 远远小于声音数据的错误率 (10^{-9})，而现成的 CD-DA 技术不能满足这一要求，因此还要采用错误校正技术。于是就开发了“黄皮书 (Yellow) 标准”。

CD-DA 存放数字化的音乐节目; CD-ROM 存放数字化的文、图、声、象等.

4.2 简述 CD-DA、HDCD、SACD 的区别

- CD-DA

CD-DA 又叫激光数字唱盘,用来存储数字音频信息,如音乐歌曲等。早期,Philips 公司、SONY 公司希望用 CD 来保存数字高保真音乐，为此制定的标准称为 Compact Disc-Digital Audio 标准，简称 CD-DA 标准（CD-Audio Book）。符合这个标准的光盘都标有“Digital Audio”的标识。正式标准定义在 1982 年发布的红皮书 (Red Book) 中，包括定义了 CD 的尺寸、物理特性、编码方式、错误校正等。

- HDCD

为改善现有 CD 记录格式的缺陷，使之既能高度兼容而在音质上又能有所突破，美国 Pacific Microsonics 公司推出了具有专利保护的 HDCD 录播新技术，它的英文全称是 High Definition Compatible Digital，译为高解析度的 CD。用 HDCD 方式编码制造的激光唱片与普通 CD 具有

高度的兼容性，用在普通的激光唱机上播放，已可领略到 HDCD 编码录音技术的优越性，如用带有 HDCD 解码功能的 CD 唱机播放，则可充分欣赏到全部释放的 HDCD 信息所特有的魅力：音质清晰细腻、动态范围广阔、信噪比极高，音色更为自然逼真。

HDCD 的编码与制造

针对传统 CD 录音格式的局限与不足，PM 公司的两位 HDCD 创始人，Keith O • Johnson 录音师和 Michael W.pflaumer 计算机专家在多年音响制作中，查找并证实了对 CD 音质影响的几个关键因素，并提出切实可行的解决方案。

HDCD 技术是在前期录音制作中即重视所录制信号的完整和精确性，采用高于常规两倍的取样频率 88.1kHz 对模拟信号进行采样，以最大限度地展宽高频响应，减少缺损性失真，高的采样率也为 HDCD 编码运算留足了空间。

用 24bit 量化其取样值为 1677216 个，它比 16bit 系统高出 256 倍，采用高位元处理技术可以提高处理精度，降低量化误差，增加动态范围至 120dB。

在模拟至数字信号转换过程中，HDCD 技术十分重视转换精度，尽量减少串音和处理的稳定性，其能够达到的指标为转换精度百万分之一，失真分量 $< -$ 120dBfs。

这个高精度、宽频带的数字信号构成 HDCD 编码制造的基础，其数据信息量十分庞大。用常规 CD PCM 编码格式无法将其容纳。如要在普通 CD 机上兼容播放，需经特殊运算编码方可。

用高采样和高比特技术进行 CD 的录音制作已被普遍认可和广泛采用，但提醒一点是目前市场上所能见到的 20、24bit CD 激光唱盘其实质应是录音过程中采用的比特数，由于 CD“红皮书”所制定的 44.1kHz/16bit 标准格式制约，这些高信息量的母带在灌制 CD 唱片时，均经过重新运算，编码制成 16bit 的 CD 唱片。因此，我们现在 CD 唱机所能解读出来的规格仍然是 16bit/44.1kHz，由于各唱片公司在转化过程所采用手法不同，我们现在能听到的不同版本的 CD 音质也的确各有千秋，但有一点可以肯定：高比特高取样技术制作的 CD 音质远胜 16bit/44.1kHz 录音格式制作的 CD。

- SACD

SACD 全称叫 Super Audio CD，是超级音频光盘系统，它是由索尼和飞利浦公司合作开发的一款具有全面取代 CD 音源实力的最新格式的数码系统。SACD 采用 DSD 数字录音技术，它的频率范围和动态范围均优于 CD。SACD 是一种新型的光盘，它不是 CD 格式，而类似 DVD 光盘，播放时需使用 SACD 专用的播放设备。

SACD 光盘结构大致与 DVD 相似，播放面有单面和双面，信息层有单层和双层。目前市场上的 SACD 光盘较多采用单面双层结构，一层是 0.6mm 基片上储存 16bits 传统 CD 格式的信号，可与 CD 兼容，另一层是 0.6mm 基片高密度的半透明层，储存 SACD 格式的信号，再将两片基片像 DVD 盘片那样粘合而成。这种光盘可以在普通 CD 播放机上播放，也可以在 SACD 播放机上播放，当然，两者的音质是有差别的。SACD 的技术指标远优于 CD，而与 DVD-Audio 相似。SACD 的核心技术是 DSD （Direct Stream Digital 直接数据流），它与 CD、DVD-Audio 的多 bit 录制原理有根本的区别。

4.3 简述在 VCD、DVD、EVD、HD-DVD 和 BlueDVD 系统中的信源编码和信道编码

	音频	视频
VCD	MPEG1 Audio	MPEG1
DVD	Dolby AC-3 DTS(D9)	MPEG2 MP@ML
EVD	EAC	MPEG2 MP@HL
HD-DVD	Dolby Digital Plus DTS++Lossy	MPEG-2, MPEG-4 AVC, VC-1
BD	DTS 5.1 Dolby Digital 5.1 7.1-channel 96/24 PCM	

5 多媒体信息分析与理解

5.1 简述常见链接分析算法及其基本思想

- HITS (Hypertext Induced Topic Selection)

描述了权威网页和中心网页之间的一种依赖关系：一个好的中心网页应该指向很多好的权威性网页，而一个好的权威性网页应该被很多好的中心性网页所指向。

基本思想：

HITS 定义了两个参数来描述网页

1. a(v) -authority，权威度描述有多少重要的网页指向它
2. h(v) -hubness，导航性描述它指向多少重要的网页

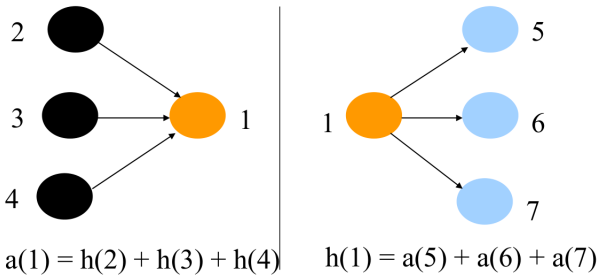

图 6: HITS

- PageRank

利用网络自身的超链接结构给所有的网页确定一个重要性的等级数，当从网页 A 链接到网页 B 时，就认为“网页 A 投了网页 B 一票”，增加了网页 B 的重要性。最后根据网页的得票数评定其重要性，以此来帮助实现排序算法的优化。

基本思想：

Page et al (1998) 提出网页的重要性取决于指向它的链接

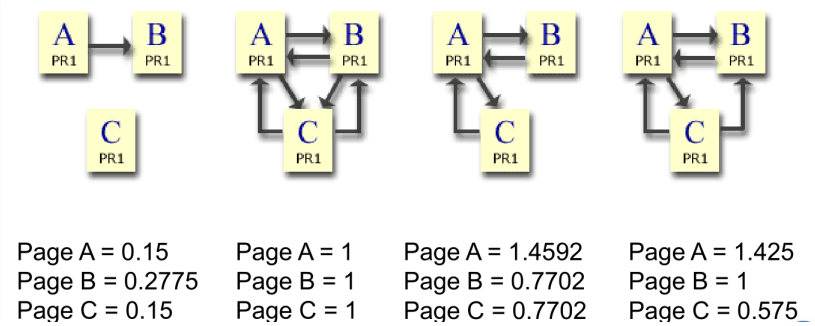

图 7: PageRank

5.2 基于内容的图像检索常用的相似度度量方法有哪些？

1. SSIM 算法—结构相似性：

   SSIM（structural similarity）是一种用来衡量图片相似度的指标，也可用来判断图片压缩后的质量。SSIM 取值范围 [0, 1]，值越大，表示图像失真越小。基本原理：SSIM 由亮度对比、对比度对比、结构对比三部分组成。亮度比较函数 l（x，y）是 x 和 y 的函数。然后，对比度比较 c（x，y）是 x 和 y 的比较。这三个组成部分相对独立。例如，亮度和/或对比度的变化不会影响图像的结构。C1、C2、C3 为常数，避免分母接近于 0 时造成的不稳定性。SSIM 函数 S 具有对称性、有界性（不超过 1）和最大值唯一性（当且仅当 x = y 时，S = 1，表示两幅图一样）。

2. Siamese network：

   Siamese Network 是一种神经网络的框架，而不是具体的某种网络，用于评估两个输入样本的相似度。两个网络分别接受输入，共享权重，然后计算两个输出向量之间的距离或者相似度，以此来判别原输入的相似性。例如判别两张脸是否为同一个人的，两个签名是否为同一个人所作。当然，siamese network 不仅只适用这种二分类问题，在目标跟踪领域也被广泛的应用如 siamMask。图中上下两个 network，都由 CNN 构成，两个模型的参数值完全相同。不同于传统 CNN 的地方，是 Siamese 网络并不直接输出类别，而是输出一组一维向量。若输入的两张图像为同一个人，则上下两个模型输出的一维向量欧氏距离较小若输入的两张图像不是同一个人，则上下两个模型输出的一维向量欧氏距离较大所以通过对上下两个模型输出的向量做欧氏距离计算，就能得到输入两幅图像的相似度

3. 均值 Hash 算法：

Hash 算法进行图片相似度识别的本质，就是将图片进行 Hash 转化，生成一组二进制数字，然后通过比较不同图片的 Hash 值距离找出相似图片。

(a) 缩小尺寸。这样做会去除图片的细节，只保留结构、明暗等基本信息，目的是统一图片大小，保证后续图片都有相同长度的哈希值，方便距离计算。网上看到的案例基本都将尺寸缩小为 8*8，64 个像素点，暂时不清楚缩小为这个尺寸的原因，但如果觉得损失的信息太多，可以将尺寸适当调大，当然像素点多了后续计算就会稍慢一些。

(b) 灰度化处理。将图片全部转换为统一的灰度图。

(c) 计算像素均值。计算像素的灰度平均值（此处均值出现）

(d) 哈希值计算。将每个像素的灰度，与平均值进行比较。大于或等于平均值，记为 1，小于平均值，记为 0，由此生成二进制数组。

(e) 图片配对，计算汉明距离。距离越近，越相似。当图片缩小为 8 * 8 时，通常认为汉明距离小于 10 的一组图片为相似图片。PS：(汉明距离：它就是将一个字符串变换成另外一个字符串所需要替换的字符个数。)

优点：速度快；缺点：精确度较差，对均值敏感

4. 差异哈希算法

5. 感知哈希算法

6. cosin 相似度

7. 皮尔逊相关系数

5.3 什么是语义鸿沟？

相似的视觉特征 (颜色、纹理、形状等) 可能表达完全不同的语义。

语义鸿沟：底层视觉特征和高层语义概念间的差距。特征相似的图像可能完全不相关，低层特征和图像意义没有必然联系。

6 实时多媒体通信

6.1 QoS 的评价参数有哪些，简述它们的基本概念

QoS(Quality of Service) 参数：

- 可用带宽

  网络的两个节点之间特定业务流的平均速率

- 时延

  数据包在网络的两个节点之间传送的平均往返时间

- 丢包率

  在网络传输过程中丢失报文的百分比

- 时延抖动

  时延的变化

- 误包率

  网络传输中报文出现错误的百分比

6.2 RTSP、MMS、RTMP、HLS 等协议完成的主要功能是什么？

- RTSP (Real Time Streaming Protocol)

由 RealNetworks 和 Netscape 提出，RTSP 通过 RTP 传送的是多媒体数据。HTTP 只能由客户机发出请求;RTSP 客户机和服务器都可以发出请求, 即 RTSP 可以是双向的。

- MMS (Microsoft Media Server Protocol)

  MMS 是微软定义的用来访问并流式接收 Windows Media 服务器中 *.asf 文件的一种协议

- RTMP(Real Time Messaging Protocol)

RTMP(Real Time Messaging Protocol) 实时消息传送协议是 Adobe 为 Flash 播放器和服务器之间音频、视频和数据传输开发的开放协议。

- HLS (HTTP Live Streaming)

HLS 点播, 基本上就是常见的分段 HTTP 点播, 不同在于, 它的分段非常小。要实现 HLS 点播, 重点在于对媒体文件分段, 目前有不少开源工具可以使用。

相对于常见的流媒体直播协议, 例如 RTMP 协议、RTSP 协议、MMS 协议等,HLS 直播最大的不同在于, 直播客户端获取到的, 并不是一个完整的数据流。HLS 协议在服务器端将直播数据流存储为连续的、很短时长的媒体文件 (MPEG-TS 格式), 而客户端则不断的下载并播放这些小文件, 因为服务器端总是会将最新的直播数据生成新的小文件, 这样客户端只要不停的按顺序播放从服务器获取到的文件, 就实现了直播。

可以认为,HLS 是以点播的技术方式来实现直播。由于数据通过 HTTP 协议传输, 所以完全不用考虑防火墙或者代理的问题, 而且分段文件的时长很短, 客户端可以很快的选择和切换码率, 以适应不同带宽条件下的播放。不过 HLS 的这种技术特点, 决定了它的延迟一般总是会高于普通的流媒体直播协议。

6.3 多媒体会议系统的基本组成与一般结构是什么？

基本组成：显示系统，音频扬声器系统，矩阵开关系统和数字，中间控制系统，环境控制系统等组成。

子系统结构：

- 数字会议系统 (发言、表决、同传等)；
- 会议场扩音系统 (会议扩音、多动能扩音)；
- 灯光舞台系统 (环境灯光、摄像灯光、舞台灯光、舞台机械)；
- 显示系统 (投影、液晶、等离子、LED，拼接、边缘融合等)；
- 摄录编播系统 (监控摄像、会议摄像、广播摄像，录像及处理)；
- 信号处理系统 (信号的分配、切换、传输、转化，图像处理等)；
- 集中控制系统 (AV 中心智能化联动控制系统)；
- 会议支撑系统 (机房、网络、远程会议、视频会议等)；
- 数字音视专业布线系统。