多媒体编码理论及进展

DUOMEITIBIANMALILUNJIJINZHAN

罗晓奔　成建斌　南丽丽　编著

中国商务出版社
CHINA COMMERCE AND TRADE PRESS

多媒体编码理论及进展

DUOMEITIBIANMALILUNJIJINZHAN

罗晓奔　成建斌　南丽丽　编著

中国商务出版社
CHINA COMMERCE AND TRADE PRESS

图书在版编目(CIP)数据

多媒体编码理论及进展/罗晓奔,成建斌,南丽丽编著. —北京:中国商务出版社,2008.12
ISBN 978-7-5103-0014-1

Ⅰ.多… Ⅱ.①罗…②成…③南… Ⅲ.多媒体技术—编码理论—研究 Ⅳ.TP37

中国版本图书馆 CIP 数据核字(2008)第 200616 号

多媒体编码理论及进展
罗晓奔 成建斌 南丽丽 编著
中国商务出版社出版
(北京市东城区安定门外大街东后巷 28 号)
邮政编码:100710
电话:010—64269744(编辑室)
010—64266119(发行部)
010—64295501
010—64263201(零售、邮购)
网址:www.cctpress.com
Email:cctp@cctpress.com

北京中商图出版物发行有限责任公司发行
三河市铭浩彩色印装有限公司印刷
787 毫米×1092 毫米 16 开本
17.625 印张 429 千字
2008 年 12 月第 1 版
2008 年 12 月第 1 次印刷

ISBN 978-7-5103-0014-1

定价:32.00 元

前　言

多媒体的应用和发展日新月异。每天我们接触到的各种信息都蕴含着极丰富的内容。而数字化技术的应用范围不仅迅速扩大到各个科学技术领域，同时也迅速渗透到工农业生产和社会生活的方方面面。因此，尽量减少信号占用的带宽、持续时间和存储容积，以节省信号在传输、处理和存储中的开销，具有巨大的经济价值。

与此同时，基于多媒体信息多半存在数据量非常大这一特点，数据压缩编码技术的发展就是一个不断进步的课题。或许，我们不应该单单把它作为一个课题来研究，实用化，更是一个需要解决的问题。而国际标准化组织对于这些编码标准的制定，为相应的编码技术的发展和交流起到了巨大的规范作用。

本书重点介绍了目前在图像、视频和音频的压缩编码方面一些主要的技术和标准，共分 8 章。第 1 章简单介绍了关于图像、视频和音频的概念，作为对后面几章的一个铺垫。第 2、3 章介绍了静止图像编码的主要技术及其标准。其中，第 2 章以 Huffman 编码、小波变换编码和零树编码为主讲述了图像编码里的一些技术。第 3 章依次介绍了静止图像编码的三个标准——JPEG 压缩编码标准、JPEG-LS 压缩编码标准和 JPEG2000 压缩编码标准。第 4、5、6 章则是围绕视频编码技术及其标准展开。其中，第 4 章介绍了视频编码的一些主要技术，如基于对象的视频编码、分布式视频编码以及流媒体和可分级视频编码等。第 5 章主要介绍了 ITU-T 的视频编码标准 H.261/H.263，以及 ISO/IEC 的视频编码标准 MPEG-X，如 MPEG-1/2/4 中的视频部分。第 6 章介绍最新的由 ITU-T 和 ISO/IEC 子委员会组成的联合视频小组 JVT 制定、公布的高级视频编码标准 H.264/AVC。这些视频编码国际标准对视频信息的压缩编码、视频通信、多媒体通信、多媒体信息存储、高清数字电视等方面的应用和发展起着巨大的推动作用，开拓了许多新的市场。第 7、8 章则是围绕音频编码技术及其标准展开。其中，第 7 章主要介绍了语音编码技术及其标准，技术方面以波形编码技术、参数编码技术和混合编码技术为主要对象，标准方面则主要介绍了语音音频标准 G.711、G.721、G.722、G.728、G.729等。第 8 章围绕 MPEG-1/2/4 中的音频部分做了介绍。

全书由罗晓奔、成建斌、南丽丽编著，并由三人负责统稿。其具体分工如下：

第 1 章，第 2 章，第 3 章，第 6 章第 2 节、第 4 节、第 5 节：罗晓奔（华南理工大学信息网络工程研究中心）；

第 4 章，第 5 章，第 6 章第 1 节：成建斌（山西交通职业技术学院信息工程系）；

第 6 章第 3 节，第 7 章，第 8 章：南丽丽（运城学院计算机科学与技术系）。

本书在写作过程中参考和引用了一些同行的研究成果、著作和论文，在此表示深深的感谢。

尽管作者多年来一直在从事这些方面的教学和科研工作，但由于多媒体压缩编码技术的飞速发展，加上作者水平有限，书中的错误和不足之处在所难免，敬请各位读者予以批评指正。

作者

2008 年 10 月

目　录

第1章 绪 论

多媒体技术是以数字技术为基础，融合通信技术（电话、传真）、广播技术（广播、电视）和计算机技术于一体，能够对文字、图形、图像、声音、视频等多种媒体信息进行存储、传送和处理的综合性高新技术。多媒体技术是人类科学技术史上继印刷术、无线电—电视技术、计算机技术之后的又一次新技术革命，在信息社会中具有十分重要的地位。多媒体技术发展前景非常广阔，从普及计算机应用、拓宽计算机处理信息的类型看，利用多媒体是计算机技术发展的必然趋势。

1.1 多媒体技术及压缩编码

随着CPU集成度和主频的迅速提高，计算机快速处理声音、图像等信息成为可能，从而出现了能对文、图、声、像等多媒体信息进行统一处理的多媒体计算机。由此可见，多媒体计算机是在计算机传统功能的基础上，扩充了处理声音、图像等信息的设备，并配备了相应的支持软件而构成的计算机系统。

多媒体技术的发展推动了多媒体计算机的性能不断提高，但多媒体技术的应用并不限于多媒体计算机。在现代生活中，从电视节目的编辑到电子出版物的制作，从VCD的播放到电视会议系统的运行，处处都包含了多媒体技术的应用。

1.1.1 多媒体技术的定义

所谓多媒体技术（multimedia computer technology）是指把文字、音频、视频、图形、图像、动画等多种媒体信息通过计算机进行数字化采集、获取、压缩和解压缩、编辑、存储等加工处理，再以单独或合成形式表现出来的一体化技术。简单地说，多媒体技术就是计算机综合处理声、文、图信息。

多媒体技术包括将媒体的各种形式转换为数字形式，以便计算机接收、存储、处理和输出。多媒体技术的研究涉及计算机的软硬件技术、计算机体系结构、数值处理技术、编辑技术、声音信号处理、图形学及图像处理、动态技术、人工智能、计算机网络和高速通信技术等很多方面。

1.1.2 压缩编码技术的应用

多媒体数据压缩和编码技术是多媒体技术中最为关键的核心技术。由于计算机只能处理数字信号（离散量信号），因此需要将模拟信号转换成计算机能够识别和处理的数字信号，即声音信息、图像信息和视频信息的数字化非常重要。数字化声音、图像和视频的数据量非常大，其存储和传输所需要的空间和时间消耗也很大，因此必须对多媒体数据进行压缩编码，才能适合多媒体信息的处理。一般情况是原始数据被压缩后存放在磁盘上或以压缩形式来传输，仅当用到它时才将数据解压缩以还原。

例如，一幅 640×480 的 256 色(8 位)图像的数据量约为 300KB，65536 色(16 位)图像的数据量约 600KB，而一分钟 CD 音质的音频文件一般需要 10MB 左右的存储空间，至于由成百上千帧彩色图像和几十分钟音频信息所组成的视频文件，其巨大的数据量更是令计算机的存储设备和数据处理能力捉襟见肘，如影像要求每秒播放 25～30 帧图像，这样，640×480 的 256 色全活动图像，要求达到每秒 7.5～9MB 的数据处理能力，而对于真彩色视频信息，则数据量将更大。因此，必须对这些多媒体信息进行数据压缩，使之适应计算机的数据处理能力和网络的数据传输速率，同时尽可能保证其视听质量不低于人们的一般接受水平。

多媒体技术中常用的数据压缩算法分为两类：无损压缩和有损压缩。无损压缩保证在数据压缩和还原过程中，多媒体信息没有任何的损耗或失真，其压缩效率通常较低；有损压缩则采用一些高效的有限失真数据压缩算法，大幅度减少多媒体中的冗余信息，其压缩效率远高于无损压缩。通常情况下，数据压缩率越高，信息的损耗或失真也越大，需要进行某种折衷，找出一个相对平衡点。这两大类数据压缩方法，又包括很多不同的算法，有着不同的应用。本书介绍的绝大多数多媒体文件格式，均采用了其中的一种或几种算法。

1.2 图像概论

图像是人类获取信息的重要来源，图像信息丰富、形象、直观、易懂，因此它是在人们日常的生活、工作中接触到最多的一类信息。在现代生活中，“图像”是一个使用频率非常高的名词，是信息时代不可缺少的内容。图像与视频紧密相关，视频是由许多许多帧图像组成的一个图像序列，视频的采集、显示都是以每帧图像为基础的，每一帧图像的质量好坏也会直接影响视频的质量。近年来，图像信息的处理和传输无论是在理论研究方面还是在实际应用方面都取得了长足的进展。对数字图像信息技术的发展、对数字视频压缩编码、处理、采集和显示起了重要的推动作用。

1.2.1 图像的分类及其表示

图像是当光辐射能量照在物体上，经过它的反射或透射，或由发光物体本身发出的光能量在人眼中所重现出的物体的视觉信息。照片、电影、电视、图画等都属于图像的范围。

1. 图像的分类

图像按其亮度等级的不同，可以分成二值图像(只有黑白两种亮度等级)和灰度图像(有多种亮度等级)两种。按其色调不同，可分为无色调的灰度(黑白)图像和有色调的彩色图像两种。按其内容的变化性质不同，有静态图像和活动图像之分。按其所占空间维数的不同，又可分为平面的二维图像和立体的三维图像等。而按其产生的机制或者捕获的方法不同，有红外图像、微波图像、CT 图像、遥感图像等。

最常见的图像是二维图像，它是三维景物在二维平面上的投影，如相机所照的图片、CCD 采集到的图像等。二维图像与平面点阵相对应，通常用矩阵来表示，矩阵的行和列值标志图像的一个点称为像素。像素值代表光线在这一点的强度，光强越强，对应像素点位置感觉越亮，像素值越大。在大多数数字图像中，像素值通常在 0～255 之间。三基色原理，一幅彩色数字

图像可以看成是对应于三个分量 R、G、B 的三个矩阵；同样，彩色数字图像也可以表示为亮度 Y 和两个色度 U、V 分量矩阵。对每个矩阵采用类似于单色图像的处理、编码方式。只是当图像以亮度和色度表示时，对 U、V 的处理可以不同于对 Y 的处理，以达到高压缩比。

2. 图像的表示

图像的亮度一般可以用多变量函数来表示：

$$I=f(x,y,z,\lambda,t)$$

其中，x,y,z 表示空间某点的坐标，t 为时间轴坐标，λ 为光的波长。当取 $z=z_0$ 时，表示二维图像；当取 $t=t_0$ 或 I 与 t 无关时，表示静态图像；当 λ 取为定值时，表示单色图像。一般来说，由于 I 表示的是物体所反射、透射或辐射的能量，因此它是正的、有界的，即

$$0\leqslant I\leqslant I_{max}$$

其中，I_{max} 表示 I 的最大值，I=0 表示绝对黑色。

f(x,y,z,λ,t)是一个复杂函数，往往根据需要将 f(x,y,z,λ,t)进行分解，一种常见的处理方法是分解为照射到物体上的光线照射分量和物体对入射光的反射分量之积，即

$$f(x,y,z,\lambda,t)=f_i(x,y,z,\lambda,t)\,f_r(x,y,z,\lambda,t)$$

式中照射分量 $0<f_i(x,y,z,\lambda,t)<\infty$，反射分量 $0\leqslant f_r(x,y,z,\lambda,t)\leqslant 1$。

当全吸收时，$f_r(x,y,z,\lambda,t)=0$；全反射时，$f_r(x,y,z,\lambda,t)=1$。其他情况取决于景物中物体的特性，反射分量包含物体的细节，与入射光无关，$f_i(x,y,z,\lambda,t)$的性质由光源来确定，与物体无关。

3. 模拟图像与数字图像

当对模拟图像信号数字化之后，模拟图像信号的具体特征就不复存在，图像信号直观的特性也体现不出，但数字化的图像却带来处理、压缩、编码、加密和传输上的极大方便，具有数字化系统的一系列特点，如高速处理能力，体积小型化，高灵活性，强抗干扰性等。视觉系统作为图像的最终归宿，仍然需要将数字处理后的图像恢复为原有形象直观的模拟图像供人眼观察。

4. 图像与图形

与图像相比，图形(Graphic)是指由人或计算机绘制出来的由许多不同笔画组成的图案，可以通过点、线、面构成复杂的图形，并且随时修改和编辑。图形通常用一些说明点、线、面特征的参数来描述，如线段的起点、终点坐标，圆心坐标和圆半径等。当然，我们可以将各类图形看作一类特殊的图像，但在一般情况下，我们提到的图像处理和图像压缩并不包括此类图形。

1.2.2 数字图像的特点

图像信息的特点可以从以下三个方面来论述：

1. 数据量大

我们知道，图像信息的内容是非常丰富的，当图像经过数字化之后，特别是彩色图像数字化之后其数据量是巨大的，例如一幅普通的黑白照片图像，若按 512×512 点阵取样，每个像素

用8bit量化，那么表示这幅图像的数据量就达512×512×8=2097152bit≈2Mbit。如果是一幅1024×1024的彩色图像，每个像素点的R、G、B分别用8bit表示，则这幅数字图像就达到1024×1024×24bit≈25Mbit的数据量。因此，海量数据是图像信息的一个最显著特点。

2. 相关性强

图像信息具有很强的多种相关性，由于自然景物、物体表面除了在边缘、轮廓、分界线之外，其变化都是缓慢的，光强变化也是平滑的。因此，数字图像的相邻行之间、相邻列之间、相邻帧之间其像素灰度变化也是相关的，而且距离越近，相关性越强。在实践中描述相关性时用的最多的是图像的相关系数，它可以直接反映任意两个像素之间的相关性，也就是在统计平均的意义上来计算它们之间的相似程度。

实验数据统计表明，在像素间隔τ为1～20个像素时，自相关系数平均值的曲线基本上呈指数规律衰减，而且相邻像素之间存在的相关性随着两者之间的距离增加而迅速减小。

3. 压缩空间比较大

当图像数字化之后，数字图像就体现不出图像的具体内容，也没有了模拟图像的直观、形象特征。尽管如此，只要在数字化过程满足采样定理，数字图像所包含的图像信息量没有减少。根据信息论原理，图像作为一个信源，描述信源的数据量是信息量(信息熵)和信息冗余量之和。

实际上，没经压缩的原始数字图像存在很大的冗余度。大量的数据是相关的，因此，数字图像在保证图像信息量的同时有很大的压缩空间。

1.2.3 图像质量评价

图像质量评价的研究是图像信息学科的基础研究之一。对于图像处理或图像通信系统，其信息的主体是图像，衡量这个系统的重要指标就是图像的质量。例如在图像编码中，就是在保持被编码图像一定质量的前提下，以尽量少的码字来表示图像，以便节省信道带宽和存储容量。

图像质量的含义包括两方面：一是图像的逼真度，即被评价图像与原标准图像的偏离程度；另一个是图像的可懂度，是指图像能向人或机器提供信息的能力。尽管最理想的情况是能够找出图像逼真度和图像可懂度的定量描述方法，以便作为评价图像和设计图像系统的依据。但是，由于目前对人的视觉系统性质还没有充分理解，对人的心理因素还找不出定量描述方法。因而用得较多、最具权威性的还是所谓主观评价方法。

(1)主观评价方法

图像的主观评价就是通过人在给定的观察条件下观察图像，对图像的优劣作主观评定，然后对评分进行统计平均，就得出评价的结果。这时，评价出的图像质量与观察者的个性及观察条件等因素有关。

(2)客观评价方法

尽管主观质量的评价是最权威的方式，但是在一些研究场合，或者由于实验条件的限制，也希望对图像质量有一个定量的客观描述。图像质量的客观评价由于着眼点不同而有多种方

法，常用的有均方误差（MSE，Mean Square Error）法、峰值信噪比（PSNR，Peak Signal Noise Ratio）法等。对彩色图像逼真度的定量表示是一个十分复杂的问题。目前应用得较多的是对黑白图像逼真度的定量表示。合理的测量方法应和主观试验结果一致，而且要求简单易行。

对于数字图像场合，设 f(i,j)为原参考图像，f(i,j)为其降质图像，逼真度可定义为归一化的均方误差（NMSE，Normalization Mean Square Error）：

$$\mathrm{NMSE}=\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\{Q[f(i,j)]-Q[\hat{f}(i,j)]\}^2}{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\{Q[f(i,j)]\}^2}$$

其中，运算符 Q[·]表示在计算逼真度前，为使测量值与主观评价的结果一致而进行的某种预处理，如对数处理、幂处理等。常用的 Q[·]为 $K_1\log_b[K_2+K_3f(i,j)]$，K_1、K_2、K_3，b 均为常数。

另外一种常用的方法为峰值均方误差（PMSE，Peak Mean Square Error）：

$$\mathrm{PMSE}=\frac{\sum_{i=0}^{N-1}\sum_{j=0}^{M-1}\{Q[f(i,j)]-Q[\hat{f}(i,j)]\}^2}{M\times N\times A^2}$$

式中，A 为 Q[f(i,j)]的最大值。实用中还常采用简单的形式 Q[f]=f。此时，对于 8bit 精度的图像，A=255，M、N 为图像尺寸。

峰值信噪比（PSNR）也是一种常用的测量方法，本质上与 PMSE 相同，其表达式为

$$\mathrm{PSNR}=10\lg\frac{255\times255}{\mathrm{PMSE}}$$

需要说明的一点是，对数字图像的评价方法仍然是一个有待进一步研究的课题。上述的三个表达式看起来直观、计算简单，但这种定量的逼真度描述和主观评价之间并没有取得真正一致性，除非对于已经达到一定显示精度、抽样比特、显示帧频等的图像。例如：对于彩色数字电视、高清晰度电视或者高码率的会议电视图像，这时两者之间的评价比较统一。但对多数情况下，逼真度的测量往往与实际观察效果不一致。这时采用的就可能是多种评价方法和测量参数，比如主观评分、PMSE 测量，有时甚至还要加上对画面的动感（帧频）评价等。

在观察一幅图像时，往往受到我们感兴趣区域的影响，图像中各感兴趣区兴趣程度的差异也影响对图像的质量评价。视觉经验告诉我们，对于给定的一幅图像，人眼对其不同区域感兴趣的程度是不同的。如头肩像，人眼往往对人脸部区域感兴趣，对其失真敏感；而对其余部位感兴趣的程度较低，允许存在较大的失真，这部分对视觉影响较小的失真却可能引起 NMSE 和 PSNR 值较大的下降，使得主、客观评价不一致。因此，为了使定量描述的客观评价更加合理，遵循人类视觉系统的视觉特性，可采用基于视觉感兴趣区域的图像质量评价方法。其基本思路是：计算各个像素点对均方误差的贡献时，考虑采用不同的加权因子，感兴趣区域的加权值大，故对均方误差的贡献大。而不感兴趣区域的加权值小，对均方误差的贡献小。

1.2.4 数字图像系统

数字图像系统是一个非常复杂的、既包括硬件又包括软件的系统，随着具体的应用目标的不同，其构成也是大不相同。但是，我们从它们最基本的功能特征出发，可以构建出一个基本

的数字图像处理系统模型。

在这个基本的系统中，包括了五大部分图像处理功能：待处理的图像信号的输入，即采集模块；已处理图像的输出，即显示模块；在处理过程中需要用到的控制和存储模块；用于传输的通信模块；最为关键的图像处理核心模块。下面简要地介绍这几个模块的组成和特点。

(1)图像输入设备

根据不同的应用需求，图像的输入设备可以采用不同的方式。如 CCD 摄像机、数码照相机、磁带录像机输出，激光视盘输出，红外、X 光摄像机、扫描仪输出等。此外，接收的广播电视信号，来自其他图像处理系统的信号等，也可以作为图像处理系统的输入。

(2)图像输出设备

目前，最常用的图像输出设备为 CRT(Cathode Ray Tube)荧光屏显示器，如电视机、计算机的显示器。平板液晶显示器(LCD，Liquid Crystal Display)和等离子显示器(PDP，Plasma Display Panel)是近年来迅速发展的一种显示设备，将很快取代相当一部分 CRT 显示器。此外，还有彩色打印机、硬拷贝机、彩色绘图仪等。

(3)图像存储和控制设备

控制设备主要是用于在图像处理过程中对主图像处理设备进行控制，如键盘、鼠标、控制杆、各种开关等；图像存储设备主要是用于在图像处理的过程中，对图像信息本身和其他相关信息进行暂时或永久的存储，如各种 RAM、ROM、闪存(Flash Memory)、硬盘、光盘、磁带机等。

(4)用户通信设备

有些情况下，用户需将已处理好的、或还要进一步处理的图像信号取出或送入主图像处理设备，该模块可满足用户的这一需求。通信相当于远端的存取操作，如基于局域网、数字通信网的通信设备等。

(5)主图像处理设备

这部分是图像处理系统的核心。主处理设备可以大到分布式计算机组、一台大型计算机，小至一台微机，甚至一片 DSP 芯片。除了硬件外，更重要的是它还包括用于图像处理的各种通用或专用软件，其规模可以是一套图像处理系统软件，也可能只是一段图像处理指令。

1.3 视频概论

在介绍过图像以及图像的一些特点之后，我们就比较容易理解视频以及视频信号的特点了。视频是由许多许多幅按时间序列构成的连续图像，每一幅图像称为一帧(Frame)。视频记录的是来自光源辐射光或场景中的反射光经平面投影后的光强度随时间变化的信号，可以认为是一个图像序列，由于每一帧的图像内容可能不同，故这个图像序列看起来就是活动图像了。例如电视信号就是一种最常见的视频信号。当人眼观看视频信号时，不仅受人类视觉系统空间频率响应的影响，需要有一定的空间分辨率，而且也受视觉系统时间频率响应的影响，需要保证图像的连续性。人类视觉系统的这种时空频率响应影响着人们对视觉质量的评价。

1.3.1 视频信号的获取和显示

视频信息是多媒体信息中最重要的一种信息，当视频信息与语音信息、图像信息、文本等有机结合在一起时构成了多媒体信息。视频信号有模拟视频信号与数字视频信号之分。当模拟视频信号数字化之后，便得到数字视频或数字序列图像，数字化是视频压缩编码的基础。一般来说，用于显示的视频信号通常都是模拟视频信号，如电视信号。而用于计算、处理的通常是数字视频信息，因为数字信号要比模拟信号更适合于计算机处理。但数字视频的数据量是非常惊人的，按照 ITU-R601(International Telecommunications Union-Radio)标准 4∶2∶2 格式的 PAL 制数字化视频信号，每帧数据量为 $720\times576\times8\times2\approx6.64$Mbit，每秒数据量为 $6.64\times25\approx165$Mbit。因此一片 600Mbit 的 CD-ROM 只能存储 727 帧左右的视频图像信号，或者说可以存储大约 29s 左右的数字电视节目。如果存储更高分辨率的高清数字电视信号，那存储时间就更短了，大约只有几秒钟时间。

1. 视频信号的获取

除了人眼感受视频信息之外，获取视频信号的最主要工具是视频摄像机，它以确定的时间间隔逐帧捕获图像。在视频摄像机里有一种类似于人眼中锥状体和杆状体一样的感光器件，我们称之为光电传感器。当摄像机只有一个光电传感器，而且对光频谱吸收函数类似于视觉系统的相对亮度效率函数时，所采集的是灰度图像。当摄像机有三个分别对三基色响应的光电传感器时，采集到的是彩色图像。如果光电传感器的频率响应范围在可见光之外，如红外线波段，X 光波段，则摄像机采集的视频信号不能被人眼所观察，而作为某种特殊用途，如夜间的红外摄影、X 光摄像等。

当视频摄像机的输出为模拟视频信号时(不经过 A/D 转换)，称为模拟摄像机，当输出为数字视频时，称为数字摄像机。由于数字化的优越性和通信环境数字化的进展，数字摄像机的出现打破了多年来模拟摄像机一统天下的局面，并且其发展的势头非常之快，数字摄像机必将逐步取代模拟摄像机。为了保证视频信号能在模拟电视上显示，数字摄像机通常继续保留模拟复合视频信号输出。

按照摄像机感光表面传感器材料的不同，通常有两种类型的摄像机：一类就是我们常见的基于半导体器件的 CCD(Charge-Coupled Devices)阵列或 CMOS 阵列摄像机；另一类是曾经广泛使用的基于视像管的摄像机，如光导摄像机、氧化铅光电导摄像机等。摄像机成像器的感光表面前放置一聚焦透镜，从景物表面反射的光线经透镜聚焦在感光表面上，感光表面上的光电传感器转换光信号到电信号。对于 CCD 摄像机，感光表面上分布着二维阵列的光电传感器，一个像素点对应一个传感器，当光线被聚焦在感光表面上时，每一个传感器都把光信号转换为电信号，并在帧间隔时间内存储在缓存中。而在读出这些信号时，则按照从上到下的顺序，一行一行地读出显示或进行处理。视像管摄像机其感光表面被随光强度变化的电子束以光栅扫描方式逐行扫描，因此，不同行扫描的时间不是同时的，这一点与 CCD 摄像机不同。对于逐行扫描，视像管摄像机中电子束连续地扫描每一行，而 CCD 摄像机连续读出缓存中的信号。对于隔行扫描，视像管摄像机中电子束在前半帧的时间里隔行扫描，在后半帧的时间里电子束扫描另外一行。而 CCD 摄像机则在前半帧时间里每隔一行读出缓存中的数据，在后半帧

时间里读出另外一行。

对于捕获彩色图像，上面所说的每一个传感器都包含三种分别对三基色频率响应的传感器。有些家用摄像机，为了降低成本每个像素点只采用一个传感器来进行彩色成像，它是通过将每个像素点区域分成三个小区域，每个小区域分别对应不同基色的频率响应。三种捕获的颜色信号可以直接以三基色形式输出或者转换成亮度和色度信号输出，也可以复合成复合信号输出。

图像通信设备的视频信号，大部分来自 CCD 模拟彩色摄像机。CCD 摄像机的体积可以很小，能在面积非常小的芯片上分布数量巨大的传感器阵列。如在 1/6 英寸的面积上能够排列 100 多万像素。由于单位面积的像素点增加，图像分辨率可以做得很高。

数字摄像机最典型的代表是 DV 标准的数字视频摄像机，它通过将 CCD 转换光信号得到的视频信号进行 A/D 转换，转换成数字视频信号，然后再经过数字信号处理、数据压缩，最终可输出经压缩的数字视频信号（如压缩比为 3∶1～5∶1）。这种数字摄像输出的图像质量较高，水平清晰度可达 540 线，已接近广播级模拟摄像机指标的下限。而家用录像系统（VHS，Video Home System）的模拟录像格式（8mm 录像带）和 S-VHS 格式的图像的水平清晰度分别为 240 线和 400 线。

DV 摄像机具有 DV/IEEE1394 输出接口（俗称“火线”，Fire Wire），是一种和 PC 机相连的接口协议。在 PC 上加上 DV 输入输出接口卡，就可以和这种数字摄像机方便地进行信息交互。当然，这类摄像通常还带有普通模拟复合视频输出及 S-Video 输出。

2. 视频信号的显示

显示视频最常见的设备是阴极射线管 CRT，或称显像管，现在使用最多的是彩色显像管。它主要由电子枪、电子束偏转系统和荧光屏组成。其中电子枪用来发射电子，并使之成为加速和聚焦的电子束，根据输入信号的大小，可以控制电子束的强弱；偏转系统使电子束作水平或垂直的偏转，使电子束根据屏幕扫描路径的要求打在荧光屏的指定位置；荧光屏随着入射电子束的强度发出不同强弱的光，从而显示出可供观看的图像。

常见的彩色显像管是单枪三束显像管，在这种显像管中，三条电子束公用一个电子枪，三条电子束水平排列，射到荧光屏上对应的像素点。由电子枪发出的三束电子流的强弱分别代表所显示像素的 RGB 三基色分量的大小。当电子流击中荧光屏某像素点上对应的 RGB 荧光粉小点时，使其发出不同的色光。一个像素的三种不同波长的色光在人眼中混合成某种颜色的光。当电子束周而复始地从左到右、从上到下快速扫描时，由于眼睛的视觉暂留作用，就会在我们眼中形成一幅幅生动的画面。

CRT 显示器显示图像具有很大的动态范围，以致显示的图像很亮，能在较远的距离或者在较亮的白天光照下观看。然而为了能使电子束到达荧光屏上下左右边上，CRT 显像管的长度大约等于屏幕的宽度。当观看的屏幕增大时，必然要求更长的显像管，使得 CRT 显示器显得笨重粗大。为了达到大屏幕、薄厚度、重量轻便的应用要求，出现了液晶显示器、彩色等离子体显示器等新显示产品。

LCD 显示器是利用液晶场致发光效应的一种新型平板显示器件。其中的液晶是一种在一定温度范围内呈现既不同于固态、液态，又不同于气态的特殊物质态，它既具有各向异性的

晶体所特有的双折射性，又具有液体的流动性。在显示应用领域，液晶由于它的各向异性而具有的电光效应，可以制成不同类型的显示器件。它的基本原理是应用薄膜晶体管阵列(TFTs，active-matrix Thin-Film Transistors)产生的电场改变液晶的光学特性，从而改变液晶的亮度和颜色。

等离子体显示器 PDP 是另一种不同于 LCD 的显示器。在 PDP 器件中，一种惰性气体(如氙气)充满在两层玻璃片之间，它间隔 100～200μm 宽平行分开排列。使用电极使气体放电产生紫外光。红、绿、蓝荧光物质吸收这些放电的紫外光的能量，再辐射出不同颜色的可见光呈现在屏幕上。因此不同于 LCD，PDP 是一种发射型显示器，它的亮度和可视场的角度这两项性能都比 LCD 要好。

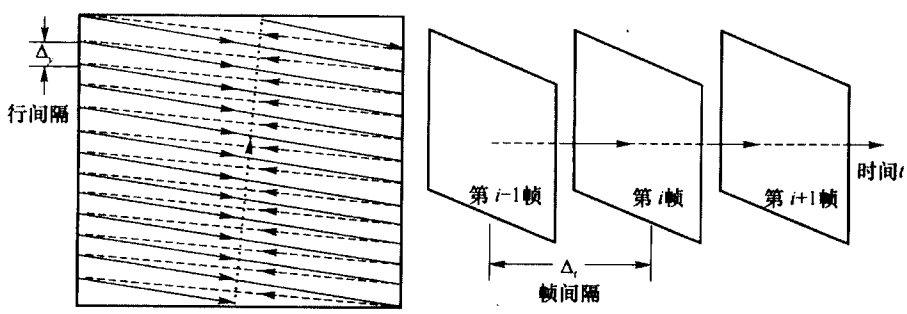
1.3.2 模拟视频和数字视频

1. 模拟视频

模拟系统使用光栅扫描的方式获取或显示视频信号。如最常见的例子是模拟黑白电视和模拟彩色电视系统。所谓光栅扫描方式是指在一定的时间间隔里光束或者电子束以从左到右、从上到下的方式扫描采集器的感光表面或者荧光屏表面，当间隔为一帧时间间隔，则捕获或显示一帧图像，当间隔为一场时间间隔，则捕获或显示一场图像。在电视系统里，两场图像为一帧。场的概念是相对于隔行扫描光栅而言的。

(1)逐行扫描(Progressive Scan)方式

光栅扫描时，摄像机是在时间上和空间垂直方向上通过采样来捕获视频信号的，以一维波形形式存储采集到的连续信号，如图 1-1(a)所示。扫描方式是以固定的行间隔 Δ_y 逐行从上至下扫描，Δ_y 也称为垂直采样间隔或行间距。时间上采样间隔是固定的帧间隔 Δ_t，每一帧由连续的水平扫描行组成。实际上扫描线稍稍有一点倾斜，因此最下面一行扫描线通常只有半行，最上面一行也是如此。为了分析简单起见，我们通常认为每一行都是水平的。视频序列就是由许多相隔为 Δ_t 的帧组成，其帧顺序如图 1-1(b)所示。



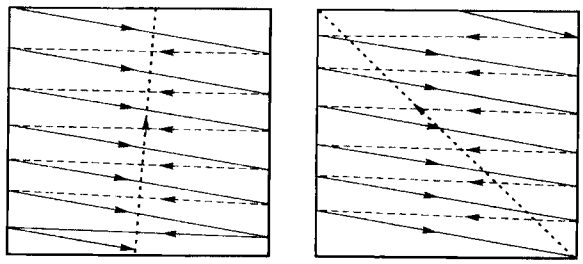
(a) 逐行扫描方式的扫描行　　(b) 逐行扫描视频帧顺序

图 1-1 逐行扫描及其帧顺序示意图

(2)隔行扫描(Interlaced Scan)方式

采用隔行光栅扫描时，每一帧分为前后两场，场间隔时间是帧间隔的一半。扫描电子束或

光束在前一场(电视信号中称为奇数场,MPEG 标准中称为顶场)时间里从左往右,从上往下每隔一行扫描一行一直到可扫描范围的最后一行,在后一场(相应地在电视信号中称为偶数场,MPEG 中称为底场)时间按照同样方式扫描其他行,如图 1-2 所示。显然,每一场图像的扫描行只有帧扫描线的一半,而每一场扫描线的垂直方向间隔是逐行扫描行间隔的两倍,将两场光栅扫描图重合在一起就构成一帧。不同系统两场的采样顺序不同,可以是顶场先采样,也可以底场先采样,关键是一帧中相邻两行采样的时间相隔相差 $\Delta_t/2$,相邻两行分别位于顶场和底场中。



(a) 隔行扫描奇数场　　(b) 隔行扫描偶数场

图 1-2　隔行扫描示意图

模拟黑白或彩色电视系统通常采用光栅扫描方式,描述光栅扫描有两个基本参数:帧频(每秒采样的帧数 fps 或用多少 Hz 表示)和每帧扫描行数,分别用字母 ω_t,ω_y 表示。这两个参数对应于光栅扫描的时间分辨率和空间垂直方向分辨率,用 H 表示图像高度,则有:

$$\omega_t = 1/\Delta_t, \omega_y = H/\Delta_y$$

Δ_y 是行间隔或垂直采样间隔,Δ_t 是帧间隔时间。注意,Δ_t 包括光束或电子束从一帧(场)的最底端回到下一帧(场)最顶端所用的帧(场)回扫时间 T_r,在隔行扫描电视系统中称为场消隐或场逆程时间。

帧频 ω_t 和每帧扫描行数 ω_y 是影响视频质量的两个重要参数。帧频影响人们观看一段视频流畅程度的感受,低于一定阈值的帧频很容易让人感觉到图像动作的跳跃和闪烁,提高帧频将使画面更加流畅和稳定。而每帧扫描行数则影响人眼对一段视频清晰程度的感受,行数越多,图像画面越清晰。这两个参数通常由人眼在不同观察环境中视觉的时空频率响应阈值决定。如电视工业界使用帧频为 25～30Hz(场频为 50～60Hz)的隔行扫描方式,电影界使用 24Hz 的帧频,而计算机界则采用大于 72Hz 的帧刷新频率等正是基于 VHS 的特性而决定的。模拟电视的行数在 500～600 行/帧,而计算机显示器则可以达到更高的扫描行数,如 VGA 的 600 行/帧,SVGA 的 1024 行/帧。计算机显示器之所以采用比电视机更高的帧频和行频,实际上就是其观看距离更近,观看的内容有更多细节(如文本和图形),空间频率(角频率)更高的缘故。

2. 数字视频

数字视频是对模拟视频信号数字化的结果,它既可以对扫描光栅采样得到,也可以直接来自数码摄像机。目前绝大多数的数码摄像机都采用 CCD 阵列器件,然后按照 CCD 传感器阵

列直接将图像场景离散化，每一个 CCD 传感器阵列元素对应一个像素，传感器阵列元素的输出对应该点光强度，因此数码摄像机本质上已经将水平和垂直方向进行离散采样了。

设 CCD 传感器阵列元素在水平方向的距离设为 Δ_x，在垂直方向距离为 Δ_y，显然对数码摄像机来说，这分别对应水平方向和垂直方向的采样间隔。而在时间域上的采样与模拟视频一样，以帧为单位，采样间隔由电视制式决定，间隔为帧间隔，$\Delta_t=1/\omega_t$。如果将扫描光栅离散化，则垂直方向采样间隔也由电视制式决定 $\Delta_y=H/\omega_y$，我们所能做的只有水平采样离散化，其间隔为 $\Delta_x=W/\omega_x$，其中 H 和 W 分别是屏幕的高度和宽度，ω_y 和 ω_x 是每帧扫描行数和每行采样点数。我们用 $\psi(m,n,k)(m,n,k=0,1,2,\cdots)$ 表示数字视频，整数 m，n 代表行和列索引，整数 k 代表帧数索引。真实的采样点位置和时间为 $x=m\Delta_x$、$y=n\Delta_y$、$t=k\Delta_t$，用 $\psi(x,y,t)$ 表示模拟视频信号。

时间和空间上的采样间隔 Δ_t、Δ_x、Δ_y 是数字视频中非常重要的参数，它保证了视频信号在时间和空间上的离散化，但是每个像素点输出即光强度仍然是模拟量，因此还需要一个用来描述每个像素输出值数字化的参数（即亮度值或三个色度值被离散的等级），用 N_b 表示，代表每个像素值量化的比特数。如黑白图像，每个像素 $N_b=8$ 表示 256 级灰度等级，彩色图像每个像素 RGB 各 8bit，因此 $N_b=24$。当彩色视频采样亮度和色度表示并且亮度和色度采用率不同时，N_b 表示的是每个采样点的等价量化比特数。如在 4∶2∶0 格式的数字视频中，每四个亮度（Y）采样点只有两个色度（U，V）采样点，于是每个像素点等价的量化比特 $N_b=(4\times8+2\times8)/4=12$bit。

由上面这些参数，还可以得出数字视频的数据率（即每秒的数据量）为

$$R=\omega_t\times\omega_y\times\omega_x\times N_b(\text{bit/s})$$

例如 PAL 制彩色视频数字化时，$\omega_t=25$（帧频每秒 25 帧），$\omega_y=576$（每帧 576 行），$\omega_x=720$（每行 720 像素点），$N_b=24$（RGB 各 8bit），则每秒的数据量为 $25\times576\times720\times24\text{bit}\approx249\text{Mbit}$。由此可知直接数字化而没有经压缩的视频数据量是十分惊人的。

为了使数字图像在监视器上正常显示图像，不至于出现扭曲（如变胖或变瘦）现象，要求像素宽高比（PAR，Pixel Aspect Ratio）与图像屏幕宽高比（IAR，Image Aspect Ratio）满足一定的关系式，即

$$PAR=IAR\times\omega_y/\omega_x$$

因为显示数字视频时，每个像素点对应一个为该像素指定颜色的小矩形区域，而矩形的宽高比即为 PAR，当 PAR 大于上式所计算的值时，显示的人会变胖；反之，则会变瘦。另一方面，为了使数字视频采样、处理更为简单，在计算机工业界，通常使用 PAR=1.0，因此在给定显示屏幕宽度和高度即 IAR 时，反过来要求每帧扫描行数和每行采样点数要满足以下要求：

$$\omega_y/\omega_x=PAR/IAR=1/IAR$$

或者说，

$$\omega_x=IAR\times\omega_y$$

例如，PAL 电视信号用于扫描图像的有效行数为 576 行，即图像在垂直方向上的分辨率为 576 点，按现行电视屏幕 IAR=4∶3 的宽高比计算，要求图像在水平方向上的分辨率应为 $576\times4/3=768$ 点，这就得到了 768×576 这一常见的 PAL 制数字图像大小。而 NTSC 制电视信号图像在垂直方向的分辨率为 480，因而水平分辨率为 $480\times4/3=640$，图像大小为 640×480 像素。