

国外电子与通信教材系列

视频处理与通信

Video Processing and Communications

Yao Wang

[美] Jörn Ostermann 著

Ya-Qin Zhang

侯正信 杨 喜 王文全 等译

PEARSON
Prentice
Hall



电子工业出版社

Publishing House of Electronics Industry
<http://www.phei.com.cn>

视频处理与通信

Video Processing and Communications

本书介绍了视频信号分析和数字视频处理的基本原理,结合国际最新科技进展对视频编码技术、国际编码标准、视频通信及网络流媒体技术进行了重点讨论,写作风格注重扎实的理论基础和复杂的系统概念之间的平衡,是目前数字视频领域的一本最具系统性、先进性和实用性的教程之一。每章末尾附有小结和习题,附录给出了有关的数学方法和部分习题答案。

作者简介

Yao Wang: 美国布鲁克林理工大学电子工程系教授。现从事视频处理和传输、多媒体信号处理、医学图像处理等相关方面的研究。1990年于加州大学圣芭芭拉分校获博士学位。1998年在美国普林斯顿大学和德国 Erlangen 大学任访问教授。1992~2000年在 AT&T 实验室任兼职研究员。2000年获纽约市长优秀青年科学技术成果奖。

Jörn Ostermann: 1994年在德国汉诺威大学获博士学位。1996年至今在 AT&T 图像处理与技术研究室从事视频编码、形状编码、多模式人机对话、标准化以及图像分析方面的研究工作。他曾任 IEEE 多媒体信号处理技术委员会会员, IEEE CAS (电路与系统) 视觉信号处理和通信技术委员会主席。1998年获得 AT&T 标准认可奖和 ISO (国际标准化组织) 奖。

Ya-Qin Zhang: 现为微软中国研究院院长。从事 MPEG2/DTV、MPEG4/VLBR 以及多媒体信息技术的研究与商品化工作。在国际会议和期刊上发表过200多篇论文,在数字视频、Internet、多媒体、无线和卫星通信方面获得40多项美国专利。1989年于美国 George Washington 大学获博士学位。他曾任 IEEE Transactions on Circuits and Systems for Video Technology 总编,是 IEEE 会士。1998年获得美国电子工程师荣誉学会授予的“杰出青年电子工程师奖”

ISBN 7-5053-7635-7



9 787505 376359 >



PEARSON
Prentice
Hall



责任编辑:赵宏英
封面设计:毛惠庚

本书贴有激光防伪标志,凡没有防伪标志者,属盗版图书
ISBN 7-5053-7635-7/TN · 1596 定价:49.00 元

国外电子与通信教材系列

视频处理与通信

Video Processing and Communications

Yao Wang

[美] Jörn Ostermann 著

Ya-Qin Zhang

侯正信 杨 事 王文全 等译

電子工業出版社

Publishing House of Electronics Industry

北京 · BEIJING

PDG

内 容 简 介

本书深入系统地阐述了视频基础理论,广泛精辟地介绍了各种实际的数字视频处理和通信系统,全面地反映了国际上数字视频技术的最新进展与成果。在概述了视频技术、视频信号分析和人类视觉系统之后,论述了数字视频信号的形成和格式转换、视频信号的数学模型、摄像机和目标的运动估计、视频编码原理和各种编码方法以及多视视频等关键课题,讨论了各种视频编码标准、视频通信中的差错控制,因特网和无线网络中的流视频等视频通信的原理和技术,每章后附有小结、习题和参考文献。

本书语言流畅、图文并茂,具有系统性、先进性和实用性特点。本书可作为理工类大专院校相关专业的研究生视频处理和通信课程的教材,也可供有关科技工作者参考或自学使用。

Simplified Chinese edition Copyright © 2003 by PEARSON EDUCATION NORTH ASIA LIMITED and Publishing House of Electronics Industry.

Video Processing and Communications by Yao Wang, Jörn Ostermann, Ya-Qin Zhang, Copyright © 2002.

All Rights Reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Prentice Hall.

This edition is authorized for sale only in the People's Republic of China (excluding the Special Administrative Region of Hong Kong and Macau).

本书中文简体字翻译版由电子工业出版社和Pearson Education培生教育出版北亚洲有限公司合作出版。未经出版者预先书面许可,不得以任何方式复制或抄袭本书的任何部分。

本书封面贴有Pearson Education培生教育出版集团激光防伪标签,无标签者不得销售。

版权贸易合同登记号:图字:01-2002-1396

图书在版编目(CIP)数据

视频处理与通信/(美)王瑶(Yao, W.)等著;侯正信等译.-北京:电子工业出版社,2003.6
(国外电子与通信教材系列)

书名原文:Video Processing and Communications

ISBN 7-5053-7635-7

I. 视... II. ①王... ②侯... III. 视频信号-信号处理-高等学校-教材, IV. TN941.1

中国版本图书馆CIP数据核字(2003)第045173号

责任编辑:赵宏英

印刷者:北京牛山世兴印刷厂

出版发行:电子工业出版社 <http://www.phei.com.cn>

北京市海淀区万寿路173信箱 邮编:100036

经 销:各地新华书店

开 本:787×1092 1/16 印张:29 字数:742千字

版 次:2003年6月第1版 2003年6月第1次印刷

定 价:49.00元

凡购买电子工业出版社的图书,如有缺损问题,请向购买书店调换。若书店售缺,请与本社发行部联系。联系电话:(010) 68279077

序

2001年7月间,电子工业出版社的领导同志邀请各高校十几位通信领域方面的老师,商量引进国外教材问题。与会同志对出版社提出的计划十分赞同,大家认为,这对我国通信事业、特别是对高等院校通信学科的教学工作会很有好处。

教材建设是高校教学建设的主要内容之一。编写、出版一本好的教材,意味着开设了一门好的课程,甚至可能预示着一个崭新学科的诞生。20世纪40年代MIT林肯实验室出版的一套28本雷达丛书,对近代电子学科、特别是对雷达技术的推动作用,就是一个很好的例子。

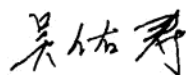
我国领导部门对教材建设一直非常重视。20世纪80年代,在原教委教材编审委员会的领导下,汇集了高等院校几百位富有教学经验的专家,编写、出版了一大批教材;很多院校还根据学校的特点和需要,陆续编写了大量的讲义和参考书。这些教材对高校的教学工作发挥了极好的作用。近年来,随着教学改革不断深入和科学技术的飞速进步,有的教材内容已比较陈旧、落后,难以适应教学的要求,特别是在电子学和通信技术发展神速、可以讲是日新月异的今天,如何适应这种情况,更是一个必须认真考虑的问题。解决这个问题,除了依靠高校的老师 and 专家撰写新的符合要求的教科书外,引进和出版一些国外优秀电子与通信教材,尤其是有选择地引进一批英文原版教材,是会有好处的。

一年多来,电子工业出版社为此做了很多工作。他们成立了一个“国外电子与通信教材系列”项目组,选派了富有经验的业务骨干负责有关工作,收集了230余种通信教材和参考书的详细资料,调来了100余种原版教材样书,依靠由20余位专家组成的出版委员会,从中精选了40多种,内容丰富,覆盖了电路理论与应用、信号与系统、数字信号处理、微电子、通信系统、电磁场与微波等方面,既可作为通信专业本科生和研究生的教学用书,也可作为有关专业人员的参考材料。此外,这批教材,有的翻译为中文,还有部分教材直接影印出版,以供教师用英语直接授课。希望这些教材的引进和出版对高校通信教学和教材改革能起一定作用。

在这里,我还要感谢参加工作的各位教授、专家、老师与参加翻译、编辑和出版的同志们。各位专家认真负责、严谨细致、不辞辛劳、不怕琐碎和精益求精的态度,充分体现了中国教育工作者和出版工作者的良好美德。

随着我国经济建设的发展和科学技术的不断进步,对高校教学工作会不断提出新的要求和希望。我想,无论如何,要做好引进国外教材的工作,一定要联系我国的实际。教材和学术专著不同,既要注意科学性、学术性,也要重视可读性,要深入浅出,便于读者自学;引进的教材要适应高校教学改革的需要,针对目前一些教材内容较为陈旧的问题,有目的地引进一些先进的和正在发展中的交叉学科的参考书;要与国内出版的教材相配套,安排好出版英文原版教材和翻译教材的比例。我们努力使这套教材能尽量满足上述要求,希望它们能放在学生们的课桌上,发挥一定的作用。

最后,预祝“国外电子与通信教材系列”项目取得成功,为我国电子与通信教学和通信产业的发展培土施肥。也恳切希望读者能对这些书籍的不足之处、特别是翻译中存在的问题,提出意见和建议,以便再版时更正。



中国工程院院士、清华大学教授
“国外电子与通信教材系列”出版委员会主任

目 录

第 1 章 视频的形成、感觉和表示	1
1.1 彩色感觉和规定	1
1.2 视频的捕捉和显示	5
1.3 模拟视频光栅	9
1.4 模拟彩色电视系统	11
1.5 数字视频	16
1.6 小结	22
1.7 习题	23
1.8 文献目录	24
第 2 章 视频信号的傅里叶分析和人类视觉系统的频率响应	25
2.1 多维连续空间信号和系统	25
2.2 多维离散空间信号和系统	27
2.3 视频信号的频域特性	28
2.4 人类视觉系统的频率响应	32
2.5 小结	37
2.6 习题	38
2.7 文献目录	39
第 3 章 视频采样	40
3.1 点阵理论基础	40
3.2 在点阵上采样	44
3.3 视频信号的采样	50
3.4 摄像机和显示器的滤波作用	57
3.5 小结	60
3.6 习题	60
3.7 文献目录	62
第 4 章 视频采样率转换	63
4.1 在不同点阵上采样信号的转换	63
4.2 视频信号的采样率转换	69
4.3 小结	80
4.4 习题	80
4.5 文献目录	83
第 5 章 视频模型	84
5.1 摄像机模型	84
5.2 照明模型	88
5.3 物体模型	91

5.4	场景模型	95
5.5	二维运动模型	98
5.6	小结	105
5.7	习题	106
5.8	文献目录	106
第 6 章	二维运动估计	108
6.1	光流	108
6.2	一般方法	111
6.3	基于像素的运动估计	117
6.4	块匹配算法	119
6.5	可变形块匹配算法	127
6.6	基于网格的运动估计	130
6.7	全局运动估计	137
6.8	基于区域的运动估计	139
6.9	多分辨率运动估计	141
6.10	运动估计在视频编码中的应用	145
6.11	小结	145
6.12	习题	146
6.13	文献目录	148
第 7 章	三维运动估计	151
7.1	基于特征的运动估计	151
7.2	直接运动估计	158
7.3	迭代运动估计	165
7.4	小结	166
7.5	习题	166
7.6	文献目录	167
第 8 章	视频编码基础	169
8.1	编码系统概述	169
8.2	概率论和信息论中的基本概念	171
8.3	信源编码的信息理论	176
8.4	二进制编码	181
8.5	标量量化	187
8.6	矢量量化	192
8.7	小结	199
8.8	习题	200
8.9	文献目录	203
第 9 章	基于波形的视频编码	204
9.1	基于块的变换编码	204
9.2	预测编码	220

9.3	采用时间预测和变换编码的视频编码	226
9.4	小结	237
9.5	习题	239
9.6	文献目录	240
第 10 章	基于内容的视频编码	242
10.1	二维形状编码	242
10.2	对于任意形状区域的纹理编码	249
10.3	形状与纹理联合编码	251
10.4	基于区域的视频编码	251
10.5	基于物体的视频编码	253
10.6	基于知识的视频编码	259
10.7	语义视频编码	260
10.8	分层编码系统	262
10.9	小结	263
10.10	习题	264
10.11	文献目录	265
第 11 章	可分级视频编码	269
11.1	可分级性的基本模式	271
11.2	基于物体的可分级性	277
11.3	基于小波变换的编码	278
11.4	小结	285
11.5	习题	285
11.6	文献目录	286
第 12 章	立体和多视序列处理	288
12.1	深度感觉	288
12.2	立体成像原理	290
12.3	位差估计	297
12.4	中间视图合成	303
12.5	立体序列编码	305
12.6	小结	308
12.7	习题	309
12.8	文献目录	310
第 13 章	视频压缩标准	312
13.1	标准化	312
13.2	采用 H.261 和 H.263 的视频电话	317
13.3	可视通信系统的标准	324
13.4	采用 MPEG-1 的消费视频通信	325
13.5	采用 MPEG-2 的数字电视	328
13.6	采用 MPEG-4 的音视频对象编码	336

13.7	视频比特流语法	350
13.8	采用 MPEG-7 的多媒体内容描述	352
13.9	小结	357
13.10	习题	358
13.11	文献目录	359
第 14 章	视频通信中的差错控制	363
14.1	动机和方法概述	363
14.2	典型的视频应用系统和通信网络	365
14.3	传输层差错控制	371
14.4	差错复原编码	374
14.5	解码器错误隐藏	381
14.6	编码器-解码器交互的差错控制	383
14.7	H.263 和 MPEG-4 中的差错复原工具	386
14.8	小结	388
14.9	习题	390
14.10	文献目录	391
第 15 章	因特网和无线 IP 网络上的流视频	397
15.1	视频流系统的结构	397
15.2	视频压缩	399
15.3	流视频的应用层 QoS 控制	399
15.4	连续媒体发布服务	404
15.5	流服务器	407
15.6	媒体同步	412
15.7	流视频协议	414
15.8	无线 IP 网络上的流视频	417
15.9	小结	422
15.10	文献目录	423
附录 A	空时梯度的确定	429
A.1	一阶和二阶梯度	429
A.2	索贝尔算子	429
A.3	高斯差分滤波器	430
附录 B	梯度下降法	432
B.1	一阶梯度下降法	432
B.2	最陡下降法	432
B.3	牛顿法	432
B.4	牛顿-拉夫森算法	433
B.5	文献目录	434
附录 C	缩写词汇表	435
附录 D	部分习题答案	442

第1章 视频的形成、感觉和表示

在第1章中,我们将叙述什么是视频信号,怎样摄取和感知它,怎样存储和传输它,以及什么是决定信号的质量和带宽(从而决定了数码率)的最重要的参数。我们首先介绍色彩感觉和再现的基础物理学(1.1节),然后叙述视频捕捉和显示的原理以及典型设备(1.2节)。大家将看到,模拟视频是以某种光栅扫描格式捕捉、存储和传输的,这种格式或者是逐行扫描或者是隔行扫描。作为例子,我们将回顾模拟彩色电视(TV)系统(1.4节),并且深入说明某些关键的参数(例如帧率和行率)是如何选择的,彩色电视信号的频谱内容是什么,信号的不同分量如何复用为一个复合信号。最后,1.5节介绍 ITU-R BT.601 视频格式(过去称 CCIR601),它是模拟彩色电视信号的数字化形式。我们给出在选择不同的数字化参数时的一些考虑。同时也介绍几个其他的数字视频格式,包括高清晰度电视(HDTV)。本章还概述了为不同的应用所开发的压缩标准及其相应的视频格式。

本章的目的是使读者得到关于模拟和数字视频的背景知识,并且了解一般的视频系统设计问题。因此,有意采取了定性的而非定量的叙述。在后面各章中,我们将回到本章所提到的某些问题并给出较严格的表述和解释。

1.1 彩色感觉和规定

视频信号是从动态的三维景物投影到视频摄像机图像平面上的一个二维图像序列。一个视频帧中任何一点的彩色值记录了在所观察的景物中一个特定的三维点所发出或反射的光。为了理解彩色值的物理意义,本节回顾一下基础的光物理学,同时对表征光及其彩色的属性加以描述。我们也要介绍人类彩色感觉的机理,以及规定彩色信号的不同方法。

1.1.1 光和彩色

光是由人眼可以感觉的、波长在 380 ~ 780 纳米(nm)范围内的电磁波组成的。光的能量是用光通量测量的,单位是瓦特,它是能量发射的速率。与我们所感觉的光的亮度直接相关的是光的辐射强度,定义为以特定方向辐射到单位立体角上的光通量,度量单位是瓦特/立体角。光源通常能发射某一波长范围内的能量,并且其强度可以在时间和空间上变化。在本书中,我们使用符号 $C(\mathbf{X}, t, \lambda)$ 表示光的辐射强度分布,它指定在波长为 λ 、空间位置 $\mathbf{X} = (X, Y, Z)$ 和时间 t 时光的亮度。

光的彩色感觉决定于其光谱成分(即它的波长组成)。例如,能量集中在 700 nm 附近的光显示红色,在整个可视频带内具有相同能量的光显示白色。一般来说,具有非常窄的带宽的光称为谱色,另一方面,白光被称为非彩色的。

有两种类型的光源:发射电磁波的照明光源和反射入射波的反射光源^①。照明光源包括太阳、灯泡、电视监视器,等等。照明光源的彩色感觉取决于它所发射能量的波长范围。照明光遵循相加规则:几个混和的照明光源的彩色感觉取决于所有光源光谱的总和。例如,红光、绿光和蓝光以正确的比例组合起来形成白色光。

反射光源是那些能反射入射光(它本身也可以是反射光)的光源。当一束光照到物体上时,在某个波长范围内的能量被吸收,而剩下的则被反射出来。反射光的彩色决定于入射光的光谱成分和被吸收的波长范围。最值得注意的反射光源是彩色染料和颜料。反射光源遵循相减规则:几种混和的反射光源的彩色感觉取决于剩余的、未被吸收的波长。例如,如果入射光是白色的,那么吸收 700 nm 附近波长的染料看起来是青色的。在这种意义上,我们说青色是红色的补色(即白色减去红色)。类似地,品色和黄色分别是绿色和蓝色的补色。青色、品色和黄色染料混和形成黑色,它吸收整个可视光谱。

1.1.2 人类的彩色感觉

人类对光的的感觉开始于位于视网膜(眼球内部的后表层)的感光细胞。有两种类型的感光细胞:锥状细胞和杆状细胞。锥状细胞在亮光下起作用,能感受彩色色调;杆状细胞工作在低环境光下,只能够提取亮度信息。来自视网膜的视觉信息通过视神经纤维传送到大脑的视觉皮层,在那里完成视觉的处理和理解。有三种类型的锥状细胞,它们在可见光谱上具有重叠的通带,其峰值分别位于红色(570 nm 附近)、绿色(535 nm 附近)和蓝色(445 nm 附近)波长,如图 1.1 所示。这些感光细胞对入射光分布的响应可用如下公式描述:

$$C_i = \int C(\lambda) a_i(\lambda) d\lambda, i = r, g, b \quad (1.1.1)$$

这里, $a_r(\lambda)$, $a_g(\lambda)$, $a_b(\lambda)$ 是指红色、绿色和蓝色锥状细胞的频率响应或相对吸收函数。这三种类型的感光细胞的组合可使人类感知任何彩色。这意味着被感知的彩色只依赖于 C_r , C_g , C_b 这三个数,而不是整个的光谱。这就是所谓彩色视觉的三感光细胞原理,是由 Young 首先提出的[17]。

描述人类彩色感觉的属性有两个:亮度和色度。亮度指被感知的光的明亮度,它是与可视频带中的总能量成正比的。色度描述被感知的光的颜色和深浅,它是由光的波长成分决定的。色度进一步由两个属性表征:色调和饱和度。色调指彩色的颜色,它是由光的峰值波长决定的,而饱和度指的是颜色有多纯,它是由光谱的范围或带宽决定的。在本书中,我们用“彩色”这个词指光的亮度和色度这两个属性,尽管通常彩色这个词只是指光的色度。

实验表明在人类视觉系统(HVS)中存在第二个处理阶段,它把锥状细胞获得的三个彩色值转换成一个与亮度成正比的值和另外两个响应于色度感觉的值,这就是所谓的 HVS 反彩色模型[2,8]。已经发现同样大小的能量在不同的波长会产生不同的亮度感觉,这种依赖于波长的亮度感觉的变化用相对亮度效率函数 $a_v(\lambda)$ 表征,也如图 1.1 所示。它本质上是所有三种类型锥状细胞频率响应的和。我们可以看到绿色波长对亮度感觉的贡献最大,其次是红色波长,而蓝色波长最小。亮度(通常用 Y 表示)与入射光谱的关系是:

^① 照明和反射光源也分别称为原发和继发光源。我们不使用这些术语是为了避免与光有关的原色相混淆。在其他资料中,照明和反射光分别被称为加色和减色。

$$Y = \int C(\lambda) a_r(\lambda) d\lambda \quad (1.1.2)$$

在上述公式中,我们忽略了时间和空间的变化,因为我们只关心在固定的空间和时间位置上的彩色和亮度感觉。我们也略去了通常与每个等式联系的比例因子,它取决于描述彩色强度和亮度的单位。

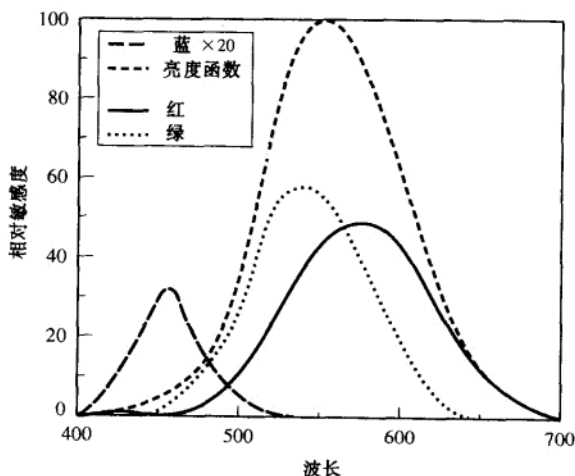


图 1.1 人类视网膜中三类锥状细胞的频率响应和亮度效率函数。图中蓝色响应曲线被放大了20倍

1.1.3 彩色混合的三基色原理

彩色物理学中一个非常重要的发现是大多数彩色可由适当选择的三种基色混和产生。这就是所谓彩色混和的三基色原理,首先由 Maxwell 在 1855 年[8,16]给出证明。令 $C_k, k=1,2,3$ 代表三种基色源的彩色, C 是一种给定的彩色。那么该原理实质上是说:

$$C = \sum_{k=1,2,3} T_k C_k \quad (1.1.3)$$

这里 T_k 是配出彩色 C 所需的三种基色的量值,称为三色激励值。一般来说,某些 T_k 可以是负的。假定只有 T_1 是负的,这意味着通过混合 C_1, C_2, C_3 不能配出 C ,但是可以用 $T_2 C_2 + T_3 C_3$ 配出 $C + |T_1| C_1$ 。实际上应该这样来选择基色,即能够通过基色正的组合产生大多数的自然彩色。照明光源最流行的基色系包括红色、绿色和蓝色,称为 RGB 基色。反射光源最通常的基色系包括青色、品色和黄色,称为 CMY 基色。事实上,RGB 和 CMY 基色系是互补的,也就是说混合一个色系中的两种彩色会产生另外一个色系中的一种彩色。例如,红色和绿色混合会产生黄色。这种互补信息可以用一个色轮给出最好的图示,这个图可在许多图像处理书籍中找到,例如参考文献[8,3]。

对于一个选定的基色系,一种确定任何彩色三色激励值的办法是首先为基色 $C_i, i=1,2,3$ 确定彩色匹配函数 $m_i(\lambda)$ 。这些函数对于整个可视频带内的各个波长 λ 描述了其谱色的三色激励值,并且可在受控观测条件下通过视觉实验确定。然后,具有光谱 $C(\lambda)$ 的任何彩色的三色激励值可由下式得到[8]:

$$T_i = \int C(\lambda) m_i(\lambda) d\lambda, i = 1, 2, 3 \quad (1.1.4)$$

为了通过正混合产生所有可视彩色,与基色相关的匹配函数必须是正的。

上述原理构成了彩色摄取和显示的基础。为了记录入射光的彩色,摄像机必须有三个传感器,其频率响应类似于所选基色系的彩色匹配函数。这可由具有期望频率响应的光学或电子滤波器来实现。类似地,为了显示一幅彩色图像,显示设备必须发出三束具有适当强度的所选基色的光束。实际上,是用电子束激发具有红色、绿色和蓝色的荧光粉。目前所有的显示系统都采用 RGB 基色,尽管为基色规定的标准光谱可能会有少许不同。同样地,彩色打印机可以通过以适当比例混合具有所选基色的三种颜料来产生不同的彩色。多数打印机采用 CMY 基色。为了更鲜艳和宽范围的彩色再现,一些彩色打印机在 CMY 基色中增加了黑色(K),从而采用 4 种基色。这就是所谓的 CMYK 基色,它能更真实地再现黑色。

1.1.4 由三色激励值规定彩色

三色激励值 彩色的激励表示规定了为产生期望的彩色所需要的三基色的比例系数,即公式 1.1.3 中的 T_k 。为了使彩色的规定独立于基色的绝对能量,这些值应该归一化,使得当 $T_k = 1, k = 1, 2, 3$ 时对应具有单位能量的基准白色(所有波长的能量相同)。当我们采用 RGB 基色时,三色激励值通常记为 R, G 和 B 。

彩色值 上述激励表示混合了彩色的亮度和色度属性。为了只测量光的色度信息(色调和饱和度),彩色坐标定义为:

$$t_k = \frac{T_k}{T_1 + T_2 + T_3}, k = 1, 2, 3 \quad (1.1.5)$$

由于 $t_1 + t_2 + t_3 = 1$, 所以两个彩色值就足以规定彩色的色度。

显而易见,图像点的彩色值依赖于采用的基色。为了使彩色的描述和规定标准化,已经定义了几种标准的基色系统。例如,彩色科学家国际组织 CIE^① 定义了一个 CIE RGB 基色系统,它包含 700(R_0), 546.1(G_0)和 435.8(B_0) nm 波长的彩色。

彩色坐标转换 我们能够把基于一种基色系的彩色值转换为基于另一种基色系的彩色值。例如,为了打印以(R, G, B)坐标存储的彩色图像,经常需要将(R, G, B)坐标转换到(C, M, Y)坐标。只要将一种基色系的激励表示用另一种基色系给出,我们就能够确定这两种坐标系之间的转换矩阵。彩色转换原理与两种基色系之间的转换矩阵推导可参见文献[8]。

1.1.5 由亮度和色度属性规定彩色

通常用于彩色显示的 RGB 基色混合了光的亮度和色度属性。在许多应用中,为了能够更高效地处理和传输彩色信号,希望将亮度与色度分量分离开来描述一种彩色。为了达到这个目的,人们开发了各种三分量彩色坐标,其中一个分量反映亮度,另外两个分量共同表示色度和饱和度。CIE XYZ 基色就是这样的一个坐标,其中 Y 直接度量亮度强度。这个坐标系中的(X, Y, Z)值与 CIE RGB 坐标系中的(R, G, B)值的关系为:

^① CIE 是 Commission Internationale de L' Eclairage 的缩写(国际照明委员会)。

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 2.365 & -0.515 & 0.005 \\ -0.897 & 1.426 & -0.014 \\ -0.468 & 0.089 & 1.009 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1.1.6)$$

除了能分离亮度和色度信息, CIE XYZ 系统的另一个优点是几乎所有的可见彩色都能由非负的激励值规定, 而这正是人们期望的特性。问题是这样定义的 X, Y, Z 彩色并不能由实际的彩色激励所实现。因此, XYZ 基色不能直接用于产生彩色, 而它主要用于定义其他的基色和彩色的数字说明。正如我们将要看到的, 用于传输彩色电视信号的彩色坐标, 例如 YIQ 和 YUV, 都是从 XYZ 坐标导出的。

也有其他的彩色表示, 除亮度外直接定义色度和饱和度。HSI 坐标就是其中之一, 这里“H”表示色度, “S”表示饱和度, “I”表示强度(等价于亮度)^①。尽管这种彩色坐标清楚地分离了光的不同属性, 但是与激励值是非线性关系并且很难计算。Gonzalez 和 Woods[3]详细介绍了各种彩色坐标及它们之间的转换。

1.2 视频的捕捉和显示

1.2.1 彩色视频成像原理

在定义了什么是光以及如何感知和表征它以后, 我们现在可以开始理解视频信号的意义了。简言之, 视频^②记录了一个观测系统(人眼或摄像机)所观测的场景中的物体发射或反射的光的强度 $C(\mathbf{X}, t, \lambda)$ 。一般地说, 该强度在时间和空间上都有变化。这里我们假定场景中有某种照明光源。否则, 将没有入射光也没有反射光, 整个图像将是黑的。当由摄像机观测场景时, 只有摄像机敏感的那些波长是可见的。令 $a_c(\lambda)$ 表示摄像机的光谱吸收函数, 那么摄像机可见的三维空间中的光强度分布为:

$$\bar{\psi}(\mathbf{X}, t) = \int_0^\infty C(\mathbf{X}, t, \lambda) a_c(\lambda) d\lambda \quad (1.2.1)$$

在任何时间 t 处, 被摄像机摄取的图像函数都是在三维场景中的光分布在二维图像平面上的投影。令 $P(\cdot)$ 表示摄像机的投影算子, 那么三维点 $\mathbf{X}$ 投影的二维位置由 $\mathbf{x} = P(\mathbf{X})$ 给出。进一步, 令 $P^{-1}(\cdot)$ 表示逆投影算子, 那么 $\mathbf{X} = P^{-1}(\mathbf{x})$ 就规定了与二维点 $\mathbf{x}$ 相应的三维位置。这样投影图像与三维图像的关系为:

$$\psi(P(\mathbf{X}), t) = \bar{\psi}(\mathbf{X}, t) \text{ 或 } \psi(\mathbf{x}, t) = \bar{\psi}(P^{-1}(\mathbf{x}), t) \quad (1.2.2)$$

函数 $\psi(\mathbf{x}, t)$ 称为视频信号。我们可以看到, 它描述了在时间 t 投影到图像平面 $\mathbf{x}$ 的三维位置 $\mathbf{X}$ 的辐射强度。通常, 视频信号具有一个有限的时间和空间范围。空间范围依赖于摄像机的观测区域, 而时间范围依赖于场景被摄取的持续时间。图像平面中的一点被称为像素(pixel, 即图像元素)或简单地记为 pel^③。对于大多数摄像系统, 投影算子 $P(\cdot)$ 可近似为一个透视投影(这个问题将在 5.1 节中详细讨论)。

① HSI 坐标也称为 HSV, 这里“V”表示强度值。

② 在本书中, 当我们提到“视频”而不加其他限时, 我们指的是一个被记录的运动图像序列。

③ 严格地讲, 符号 pixel 或 pel 只定义在数字图像学中, 在那里视频记录中的每一幅图像或每一帧由一个有限的二维像素矩阵表示。

如果摄像机吸收函数等同于人类的相对亮度效率函数,即 $a_c(\lambda) = a_v(\lambda)$,则形成亮度图像。如果吸收函数在一个窄带中是非零的,则形成单色(或单色调)图像。为了感知所有的可见彩色,根据三基色的视觉原理(见 1.1.2 节),需要三个频率响应类似于所选基色的彩色匹配函数的传感器。如前所述,多数彩色摄像机利用红色、绿色和蓝色传感器来获得彩色。

如果摄像机只有一个亮度传感器,那么 $\psi(\mathbf{x}, t)$ 是表示投影光亮度的标量函数。在本书中,我们用灰度级这个术语来表示这样一幅图像,术语黑和白将严格地用于描述只有两种彩色黑和白的图像。另一方面,如果摄像机有三个分离的传感器,每一个接收一种所选的基色,那么信号就是一个在每个点包含三个彩色值的矢量函数。我们可以利用前面章节中所介绍的其他的彩色坐标(每个包含三个值)来表征光,而不是直接规定这些彩色值。

注意,有时为了特殊的目的,我们可以采用频率不在人眼可见范围内的传感器。例如在 X 光成像中,传感器要对 X 光的光谱范围很敏感。另一方面,红外线摄像机对红外区域敏感,能在很低的环境光下工作。这种摄像机能够“看到”人眼不能感觉的物体。另一个例子是测距摄像机,其传感器发出激光束并测量它到达物体和返回传感器所需的时间。因为这个来回的时间正比于传感器到物体表面的距离,所以测距图像内任何一点的图像亮度就描述了相应的三维点到摄像机的距离或射程。

1.2.2 视频摄像机

目前所有的模拟摄像机都是一帧一帧地捕捉视频,帧与帧之间有一定的时间间隔。一些摄像机(例如电视摄像机和消费类的便携式视频摄像机)通过扫描具有一定行间隔的相继的行来获得一帧。类似地,所有显示设备都以相继的一系列的帧显示视频;对于电视监视器,扫描线相继地回扫形成分离的行。这种摄取和显示机制是根据如下事实设计的,即 HVS 不能感知很高频率的时间和空间的变化(HVS 的这个性质将在 2.4 节中更深入地讨论)。

有两种基本类型的摄像机:(1)基于光电管的摄像机,例如光导摄像管、氧化铅摄像管或正析摄像管;(2)固态传感器,例如电荷耦合器件(CCD)。摄像机镜头把场景中的图像聚焦到摄像机析像管的光敏表面,由它将光信号转换成电信号。析像管的光敏表面一般是用电子束或其他的电子方法一行一行地扫描(称为光栅扫描),然后每一帧中的扫描行被转换成用不同电压代表不同光强度的电子信号。因此不同的扫描行是以相继的方式在略微不同的时间上摄取的。对于逐行扫描,电子束是相继地扫描每一行;而对于隔行扫描,电子束是在半帧(一场)的时间隔一行扫描一行,然后扫描另外一半的行(我们将在 1.3 节详细讨论光栅扫描。)对于 CCD 摄像机,光敏表面由二维传感器矩阵组成,每个传感器对应一个像素,到达每个传感器的光信号被转换成一个电信号。在每帧时间内摄取的传感器值首先存储在缓冲器中,然后一次一行地相继读出,以形成光栅信号。与基于光电管的摄像机不同,同一帧中所有读出的值是同时被摄取的。对于隔行扫描的摄像机,每场中的扫描行是被交替读出的。

为了摄取彩色,通常有三种类型的光敏表面或 CCD 传感器,每个传感器的频率响应决定于所选基色的彩色匹配函数,如 1.1.3 节所述。为了减少成本,多数消费类摄像机采用单片 CCD 进行彩色成像。这是通过把每个像素的传感器区域分成 3 个或 4 个子区域,每个子区域感应于不同的基色实现的。摄取三个彩色信号可以转换成一个亮度信号和两个色度信号,并作为分量彩色图像发送出去,或者复用成一个复合信号。这方面的内容将在 1.2.4 节进一步解释。

现在许多摄像机是基于 CCD 的,因为它们为获得同样的空间分辨率可以比基于光电管的摄像机做得更小和更轻。CCD 技术的优点使它能够以很小的芯片摄取很高分辨率的图像矩阵。例如,消费者使用的便携式摄像机通常采用 38 万像素的 1/3 英寸的 CCD,而 200 万像素的 2/3 英寸的 CCD 已经开发用于 HDTV。基于光电管的摄像机较大且较昂贵,只用于一些特殊的应用场合,例如需要很高分辨率或需要在低环境光下有高敏感度的场合。除了彩色成像电路,多数摄像机还能实现彩色坐标转换(从 RGB 转换到亮度和色度)以及复合亮度和色度信号。为进行数字输出,也包含有模拟到数字(A/D)转换。包括在专业视频摄像机中的这种典型处理如图 1.2 所示。该摄像机提供了模拟和数字两种输出模式,而模拟输出又包括分量(CN)和复合(CS)两种格式。为了提高图像质量,该摄像机引进了数字处理技术。用于三个 CCD 中每一个 CCD 的 A/D 和预处理(包括点和区域的处理,数据检测和校正)都工作在 CCD 的输出速率 $f_{s,1}$ 。在它后面是以 2 倍 CCD 输出速率进行的图像增强和非线性滤波。为了匹配标准摄像机的输出速率 $f_{s,2}$ (对于 1.5.2 节讨论的 ITU-R BT.601 数字视频 $f_{s,2} = 13.5 \text{ MHz}$),需要 $2f_{s,1}$ 到 $f_{s,2}$ 的速率转换。关于视频摄像机和显示技术的详细说明见文献[5]。

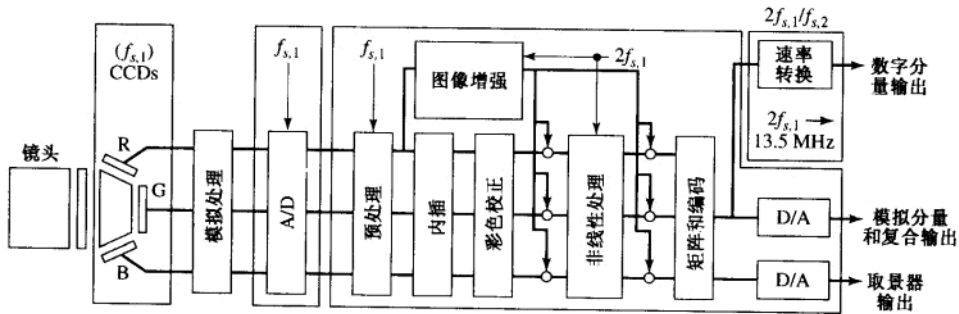


图 1.2 专业彩色视频摄像机方块示意图

1.2.3 视频显示

为显示视频,最普通的设备是阴极射线管(CRT)。在 CRT 监视器中,电子枪一行一行地向屏幕发射电子束以激励荧光粉,其强度正比于在相应位置的视频信号的强度。为了显示彩色图像,三个电子枪发射三个电子束,在每个位置以期望的强度组合激励红色、绿色和蓝色荧光粉。为了更精确,每个彩色像素由排列成小三角形的三个元素组成,称之为三元组。

CRT 能产生有很大动态范围的图像,因此显示的图像可以很亮,足以在白天或在远处观看。然而,为使电子到达屏幕的边界,CRT 的深度需要大致与屏幕的宽度相当。这样大屏幕监视器就太庞大了,不适用于应用在小形轻便的设备中。为克服这个问题,人们开发了各种不同的平板显示器。一种流行的设备是液晶显示器(LCD)。LCD 的主要思想是通过施加电场改变光学特性进而改变液晶的亮度或彩色。电场由一个晶体管阵列产生或刷新,例如在 LCD 中采用驱动矩阵薄膜晶体管(TFT)或采用等离子。等离子技术消除了对薄膜晶体管的需求,从而使大屏幕 LCD 成为可能。也有平面 CRT 的新设计。关于视频显示技术更详细的描述可参见参考文献[5]。

刚刚介绍的光栅扫描和显示机制只应用于视频摄像机和显示器。对于运动图像摄像机,

摄像机在任何帧瞬间摄取的彩色图案全部记录在胶片上。显示时,用模拟光学投影系统回放相继记录的帧。

1.2.4 复合视频与分量视频

理想情况下,无论是三色激励表示还是亮度色度表示,彩色视频都应该由三个函数或信号规定,每个描述一个彩色分量。这种格式的视频称之为分量视频。主要是由于历史的原因,也存在各种复合视频格式,其中的三个彩色信号被复用成一个单独的信号。当彩色电视系统首次开发出来时发明了这种复合格式,当时要求彩色电视信号的传输方式能使黑白电视机从中抽取亮度分量。构造复合信号依赖于这样一个性质,即色度信号具有比亮度分量小得多的带宽。通过将每个色度分量调制到一个位于亮度分量高端的频率上,并把已调色度分量加到原始亮度信号,就产生了一个包含亮度和色度信息的复合信号。为了在彩色监视器上显示复合视频信号,用滤波器把已调色度信号从亮度信号中分离出来。然后把产生的亮度和色度信号转换成红色、绿色和蓝色分量。对于灰度级监视器,只提取亮度信号并直接显示。

现在所有模拟电视系统都以复合格式传输彩色电视信号。复合格式也用于将视频存储在某种模拟磁带上(例如 VHS)。除了与灰度级信号兼容以外,复合格式消除了处理彩色信号时使不同的彩色分量同步的需要。复合信号的带宽比三个分量信号带宽的总和小,因此能被更有效地传输或存储。然而这些优点是以降低图像质量为代价的;经常由于色度和亮度分量的串扰而形成可觉察的人工痕迹。

作为数码率与图像质量之间的一个折中,发明了 S-video,它由两个分量组成:亮度分量和由两个原始色度信号复合成的一个单独的色度分量。许多高级消费类视频摄像机和显示设备都能以 S-video 格式记录或显示视频。分量格式只用于专业的视频设备。

1.2.5 伽马校正

我们说过由摄像机摄取的视频帧反映了成像景物的彩色值。事实上,多数摄像机输出的信号并不是与实际的彩色值成线性关系,而是一种非线性的形式^①:

$$v_c = B_c \gamma_c \quad (1.2.3)$$

这里 B_c 表示实际的光亮度, v_c 是摄像机的输出电压。 γ_c 的取值范围从 1.0(对于多数 CCD 摄像机)到 1.7(对于光导管摄像机)[6]。类似地,大多数显示设备的输入电压值与显示的彩色强度之间也有这种非线性的关系,即:

$$B_d = v_d \gamma_d \quad (1.2.4)$$

对于 CRT 显示器,典型的 γ_d 是 2.2~2.5[6]。为了呈现真实的彩色,我们必须在摄像机输出端加入一个相反的幂函数。类似地,在发送要显示的真实图像值之前,我们必须对显示设备的“伽马效应”进行预补偿。这些过程被称为伽马校正。

理想情况下的电视广播,在电视广播的发送端,被电视摄像机摄取的 RGB 值应该首先基于摄像机的伽马值进行校正,然后转换到用于传输的彩色坐标(对于 NTSC 是 YIQ,对于 PAL 和 SECAM 是 YUV)。在接收端,收到的 YIQ 或 YUV 值应该首先被转换成 RGB 值,然后用监视器

^① 更精确的关系是 $B_c = \alpha_c \gamma_c + B_0$, 这里 α_c 是增益因子,而 B_0 是光强截止电平。当我们假定输出电压值经过了适当的平移和缩放时,则所给出的公式是正确的。