



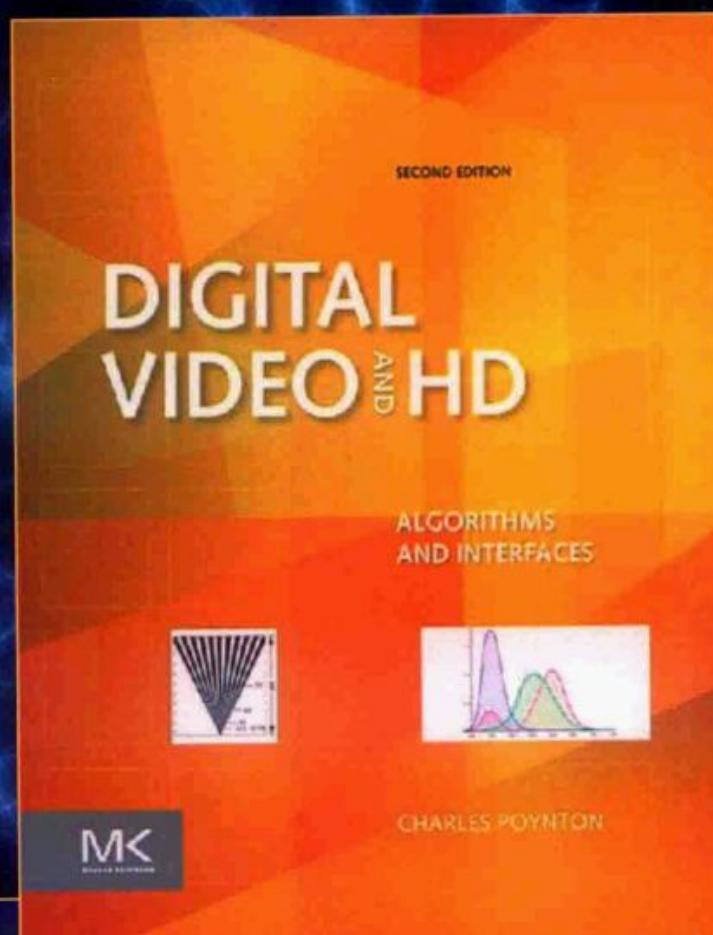
数字视频和高清 算法和接口

(原书第2版)

[加] 查尔斯·波因顿 (Charles Poynton) 著

刘开华 褚晶辉 马永涛 吕卫 宫霄霖 等译

Digital Video and HD
Algorithms and Interfaces
Second Edition



机械工业出版社
China Machine Press

图书在版编目 (CIP) 数据

数字视频和高清：算法和接口（原书第2版）/（加）查尔斯·波因顿（Charles Poynton）著；
刘升华等译。—北京：机械工业出版社，2017.5
(国外电子与电气工程技术丛书)

书名原文：Digital Video and HD: Algorithms and Interfaces, Second Edition

ISBN 978-7-111-56650-2

I. 数… II. ①查… ②刘… III. 数字视频系统 IV. TN941.3

中国版本图书馆 CIP 数据核字（2017）第 086854 号

本书版权登记号：图字：01-2013-6485

Digital Video and HD: Algorithms and Interfaces, Second Edition

Charles Poynton

ISBN: 978-0-12-391926-7

Copyright © 2012 by Elsevier Inc. All rights reserved.

Authorized Simplified Chinese translation edition published by the Elsevier (Singapore) Pte Ltd.

Copyright © 2017 by Elsevier (Singapore) Pte Ltd. All rights reserved.

Printed in China by China Machine Press under special arrangement with Elsevier (Singapore) Pte Ltd. This edition is authorized for sale in China only, excluding Hong Kong SAR, Macau SAR and Taiwan. Unauthorized export of this edition is a violation of the Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书简体中文版由 Elsevier (Singapore) Pte Ltd. 授权机械工业出版社在中华人民共和国境内（不包括香港、澳门特别行政区及台湾地区）出版及标价销售。未经许可之出口，视为违反著作权法，将受法律之制裁。

本书封底贴有 Elsevier 防伪标签，无标签者不得销售。

本书从基本概念、基础理论、实际问题、演播室标准、视频压缩、分发标准 6 个大的方面系统阐述了数字视频与高清技术。书中对基本概念的诠释极为精准，这在同类书籍中较为罕见。

本书适合作为高等院校计算机、电子与信息工程、通信工程、数字媒体等相关专业高年级本科生和研究生的“数字视频技术”课程教材或教学参考书，也可供在视频技术、图像处理与通信、电视工程技术、影视技术、多媒体技术、电影技术等领域从事研究、开发、生产及维护的科技人员阅读、参考。

出版发行：机械工业出版社（北京市西城区百万庄大街 22 号 邮政编码：100037）

责任编辑：张梦玲

责任校对：殷 虹

印 刷：北京诚信伟业印刷有限公司

版 次：2017 年 6 月第 1 版第 1 次印刷

开 本：185mm×260mm 1/16

印 张：22.75（含 0.75 印张彩插）

书 号：ISBN 978-7-111-56650-2

定 价：99.00 元

凡购本书，如有缺页、倒页、脱页，由本社发行部调换

客服热线：(010) 88378991 88361066

投稿热线：(010) 88379604

购书热线：(010) 68326294 88379649 68995259

读者信箱：hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问：北京大成律师事务所 韩光 / 邹晓东



图 2-2 作为一种戏剧化设备的色彩

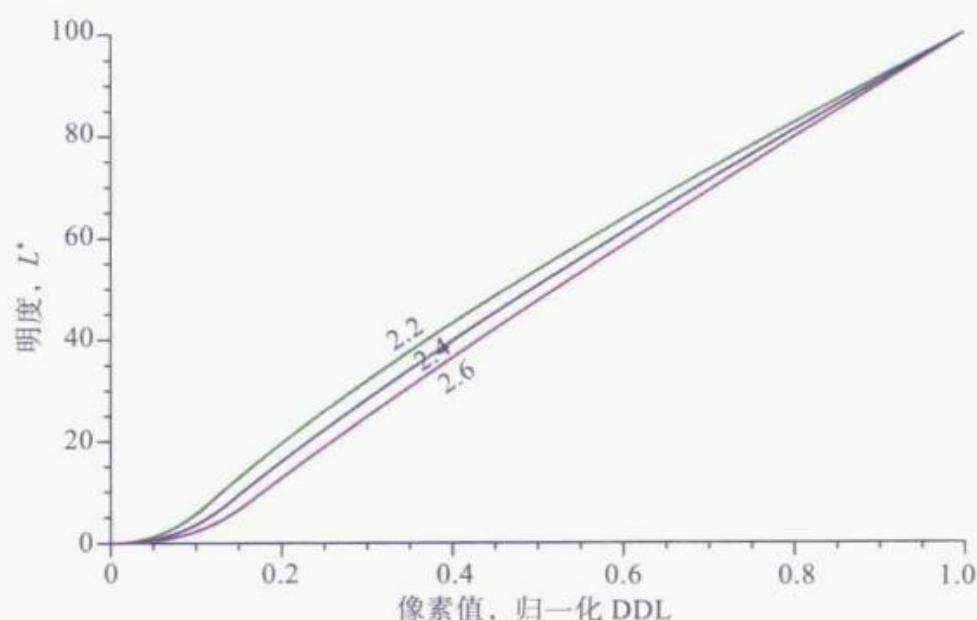


图 5-1 像素值和 L^* 的关系

在 x 轴绘制归一化像素（或者视频信号）值。相应的 L^* 值绘制为 2.2- 罩函数（典型的计算机图像，例如 sRGB 标准），2.4- 罩函数（典型的演播室视频），2.6- 罩函数（标准的数字电影）的显示罩函数（“伽马”罩函数）。在该图中，以及接下来的，当 $R=G=B$ 的适用于三个信道，或者适当的缩放（其中，另外两个信道为 0）的单个通道

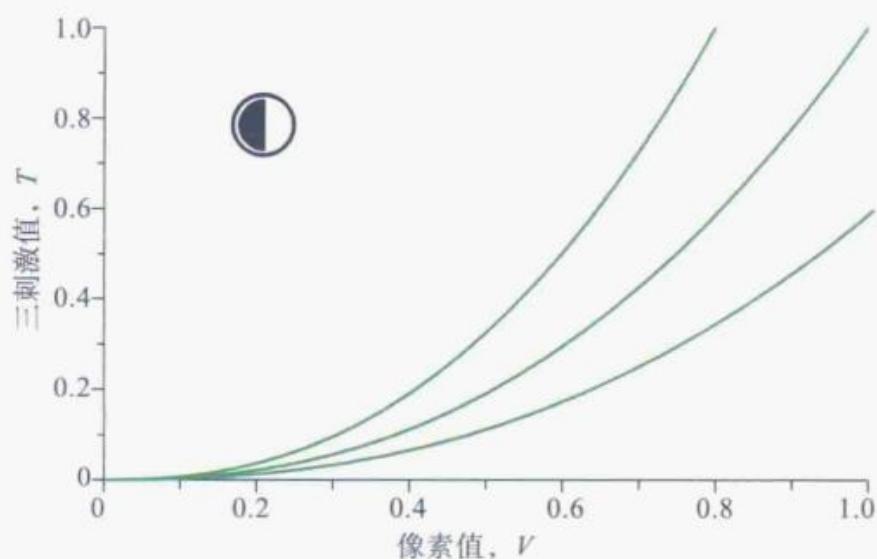


图 5-2 控制增益的影响

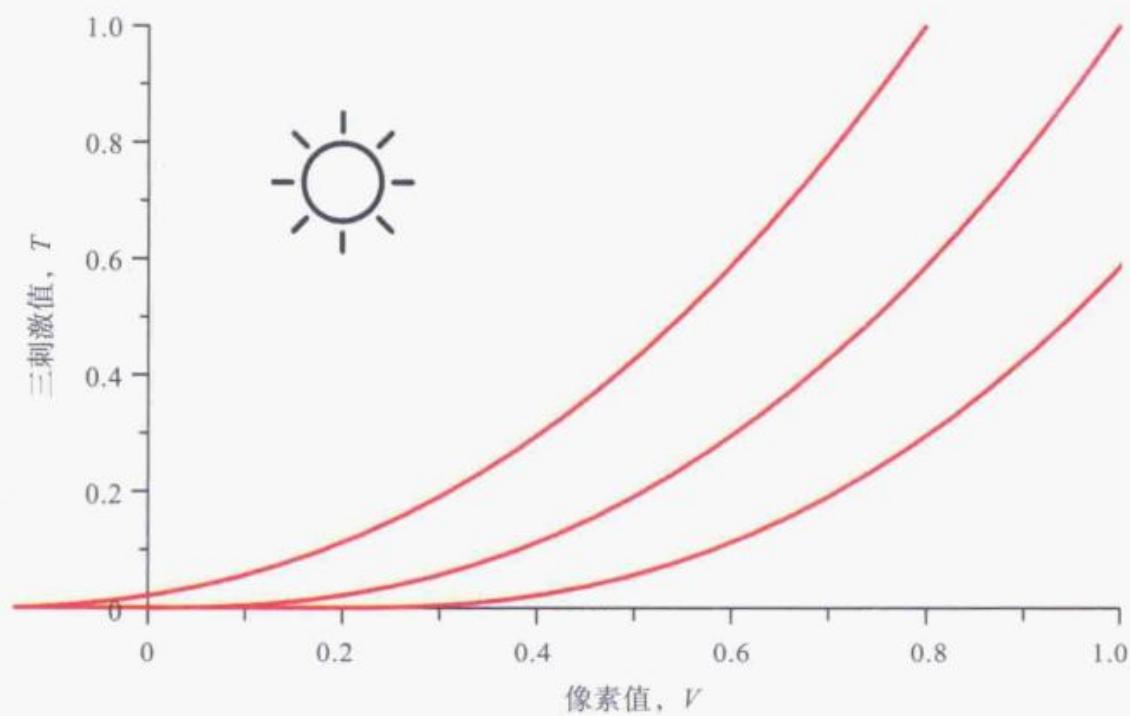


图 5-3 偏移控制的影响

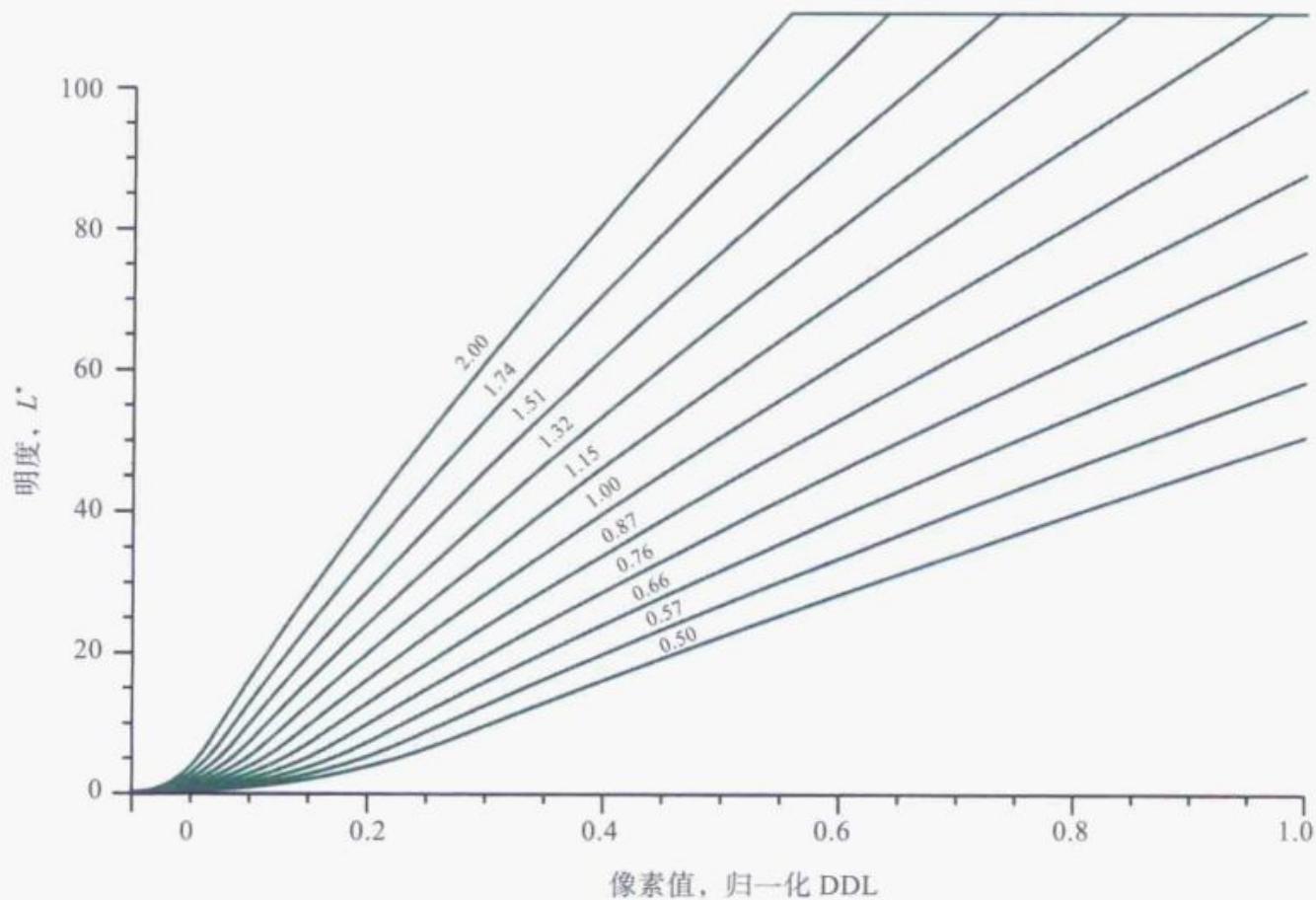


图 5-4 增益控制效果

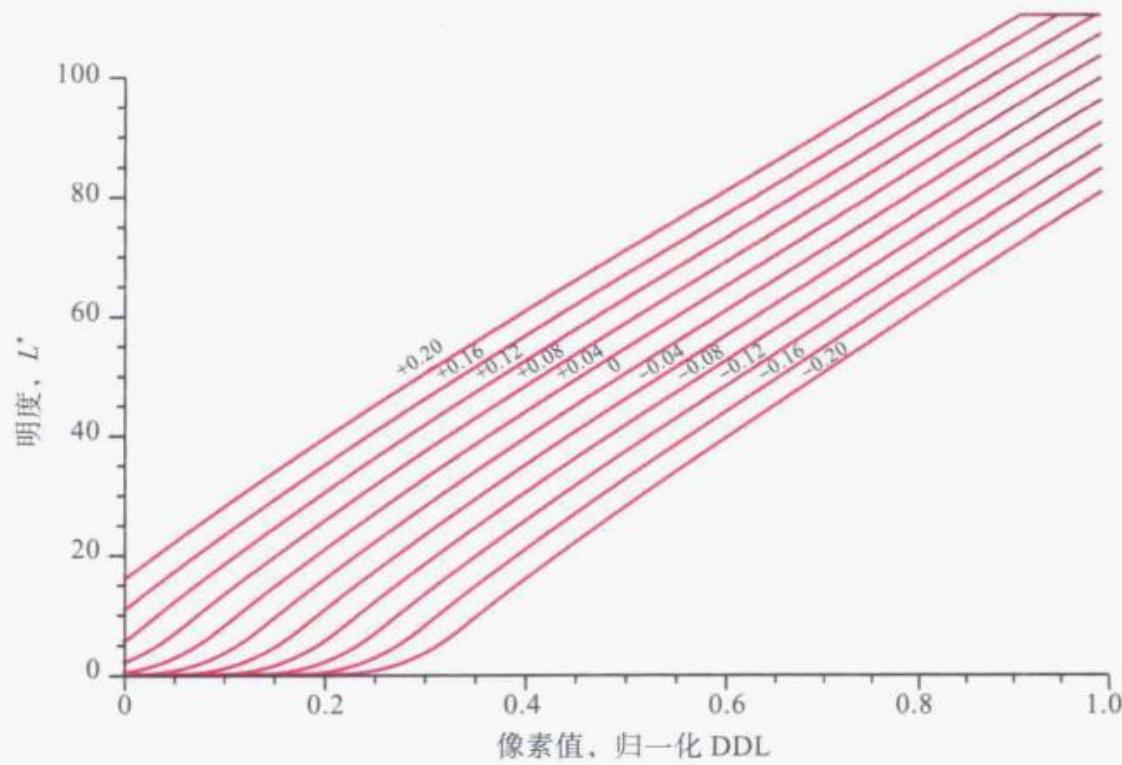


图 5-5 偏移控制效果

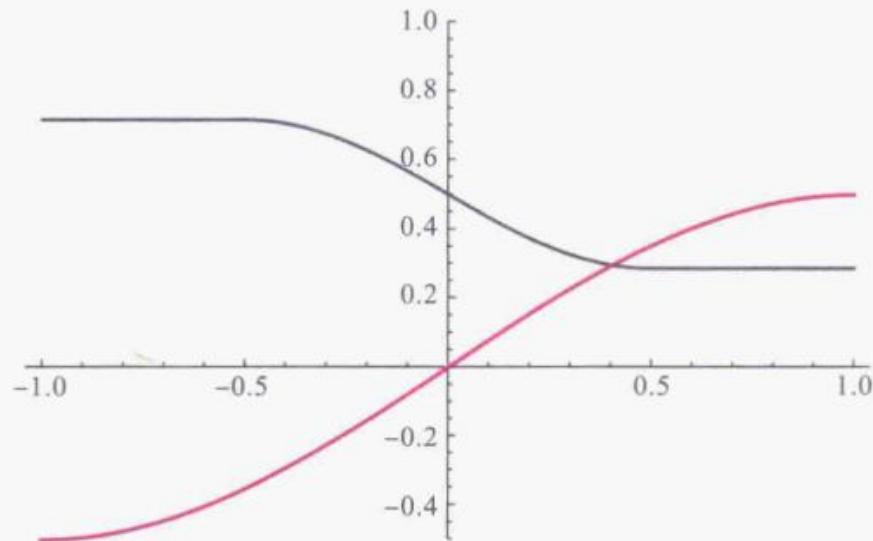


图 10-12 SD 色条中绿色 – 洋红颜色过渡段 Y' 和 C_B/C_R 的波形

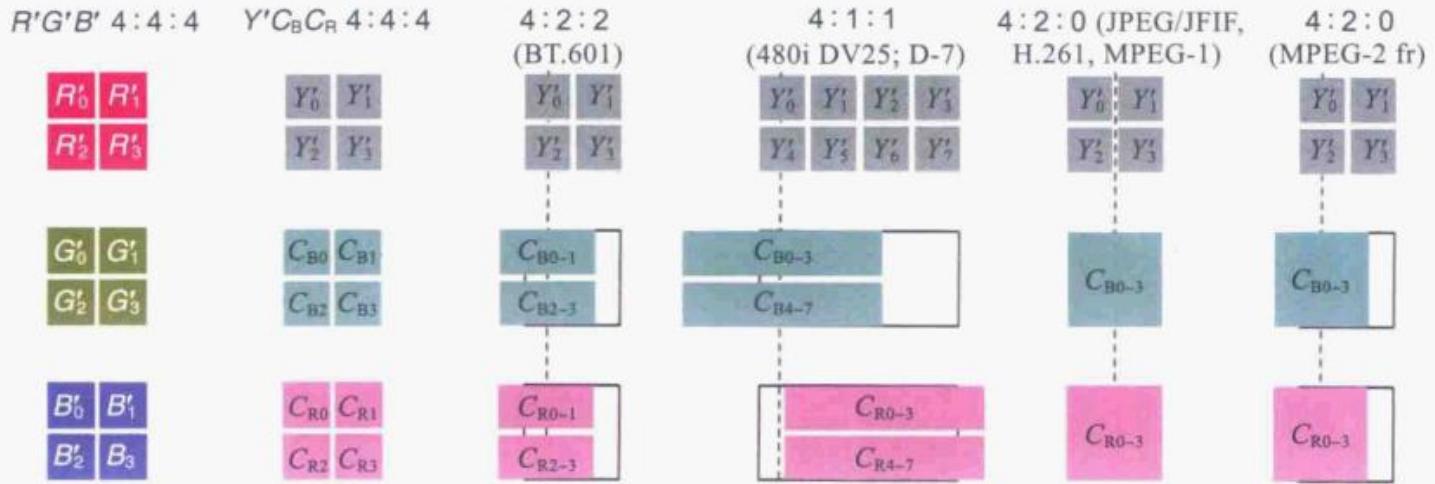


图 12-1 色度亚采样

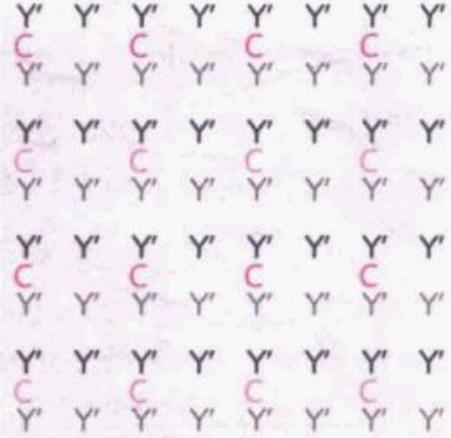
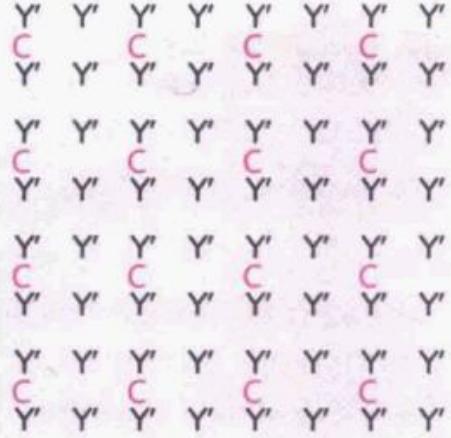
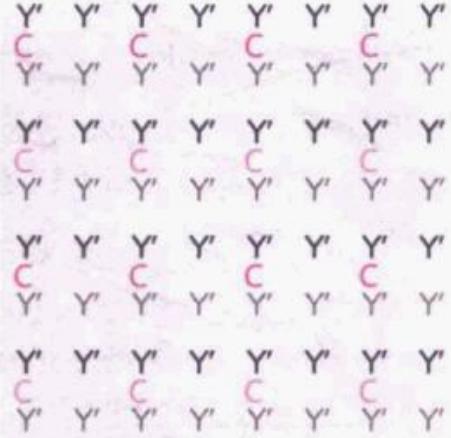
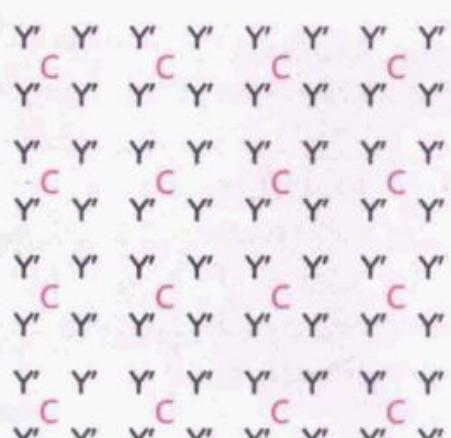
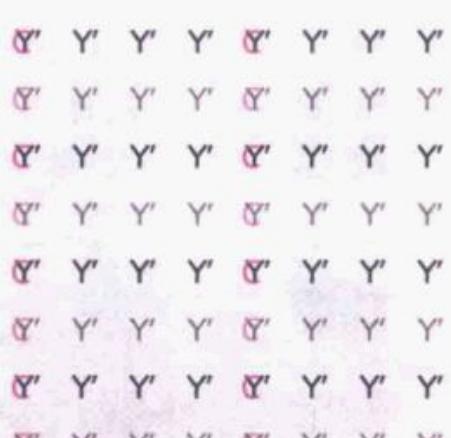
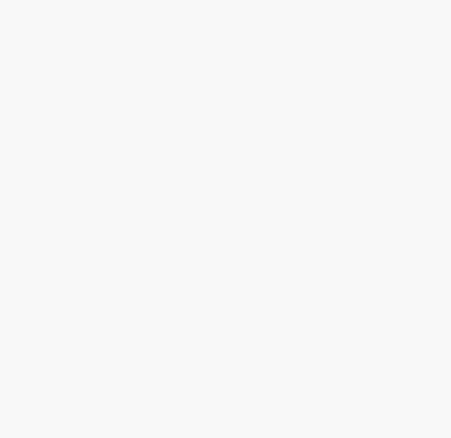
逐行								4 : 2 : 2	隔行							
																
								(BT.601)								
逐行																
								MPEG-2								
逐行																
逐行																
								JPEG/JFIF								
(逐行)																
逐行																
																

图 12-2 亚采样方案总结

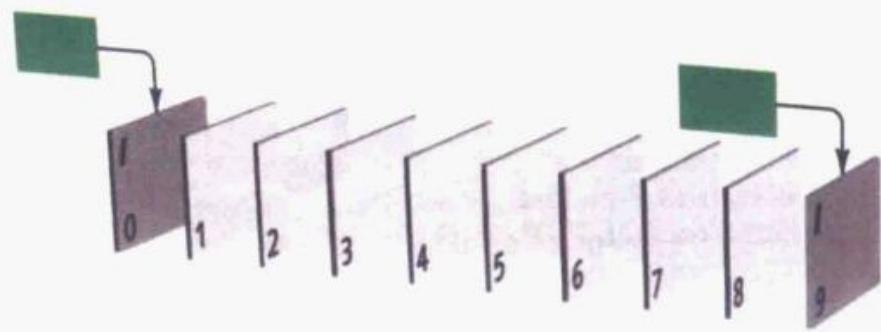


图 16-2 MPEG 图像组 (GoP)

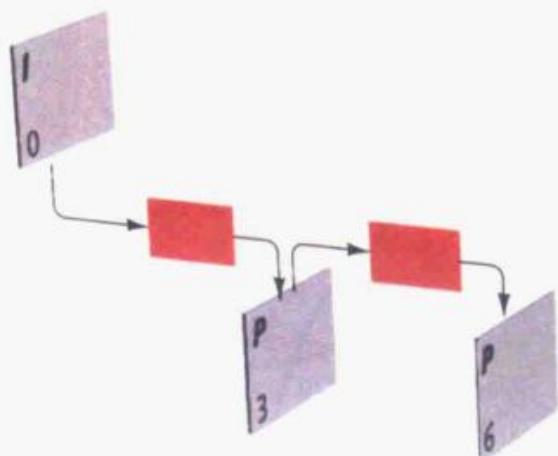


图 16-3 MPEG P 图像

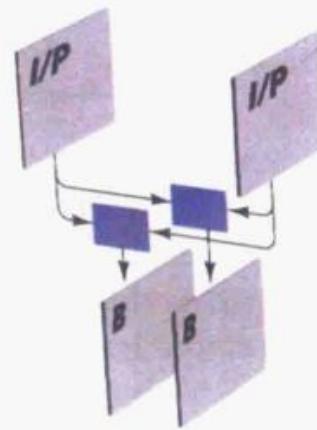


图 16-4 MPEG B 图像

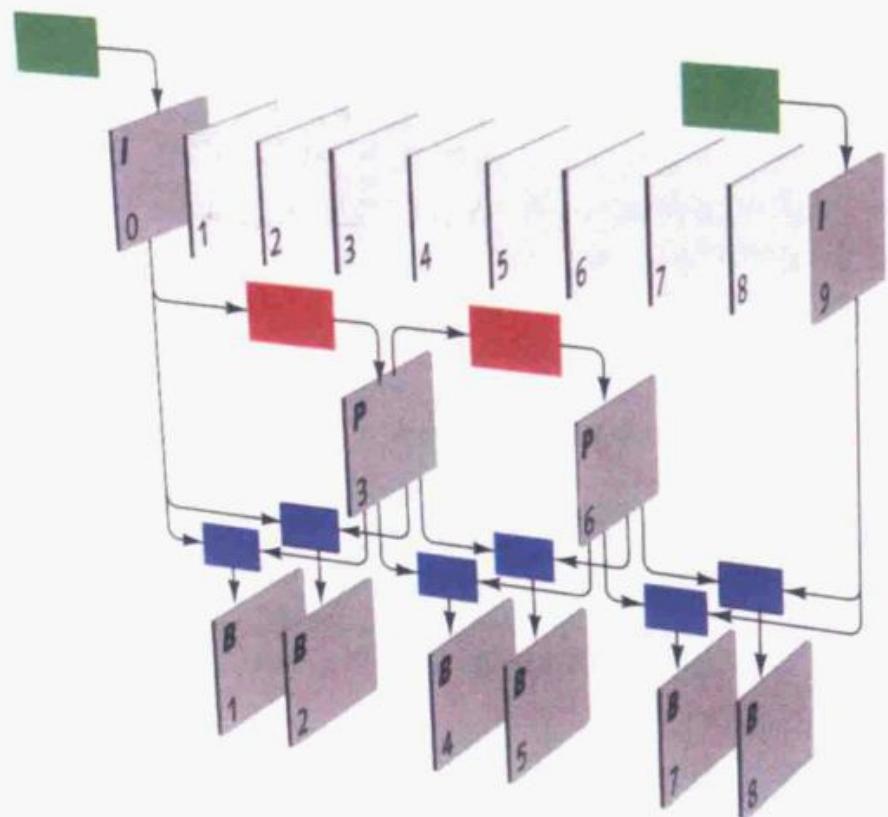


图 16-5 三级 MPEG 图像层次结构

$I_0 B_1 B_2 P_3 B_4 B_5 P_6 B_7 B_8$

图 16-6 GoP 示例

$I_0 P_1 P_2 P_3 P_4 P_5 P_6 P_7 P_8$

图 16-7 无 B 图像 9 图像 GoP 示例

$I_0 P_3 B_1 B_2 P_6 B_4 B_5 (I_9) B_7 B_8$

图 16-8 为传输重排序的 GoP

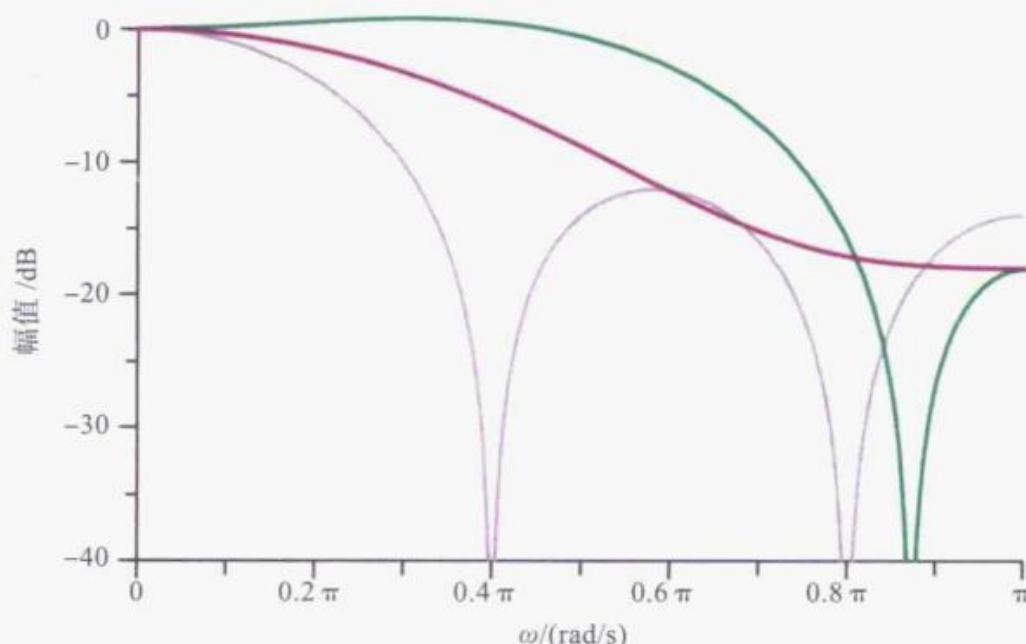


图 20-17 5 抽头 FIR 滤波器的响应

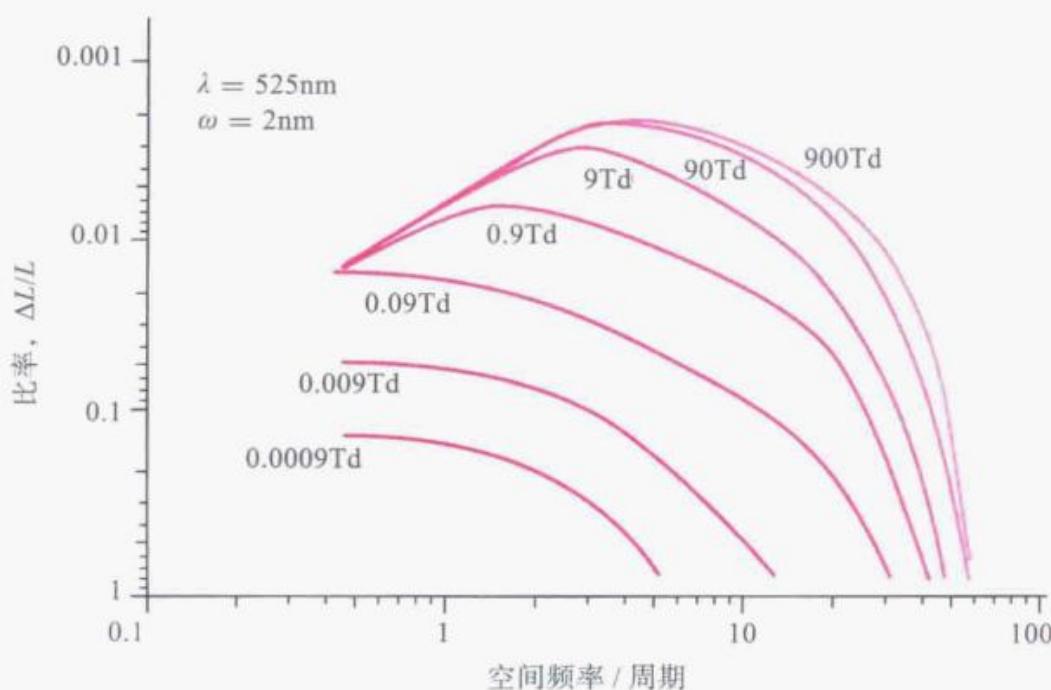


图 23-5 对比灵敏度函数

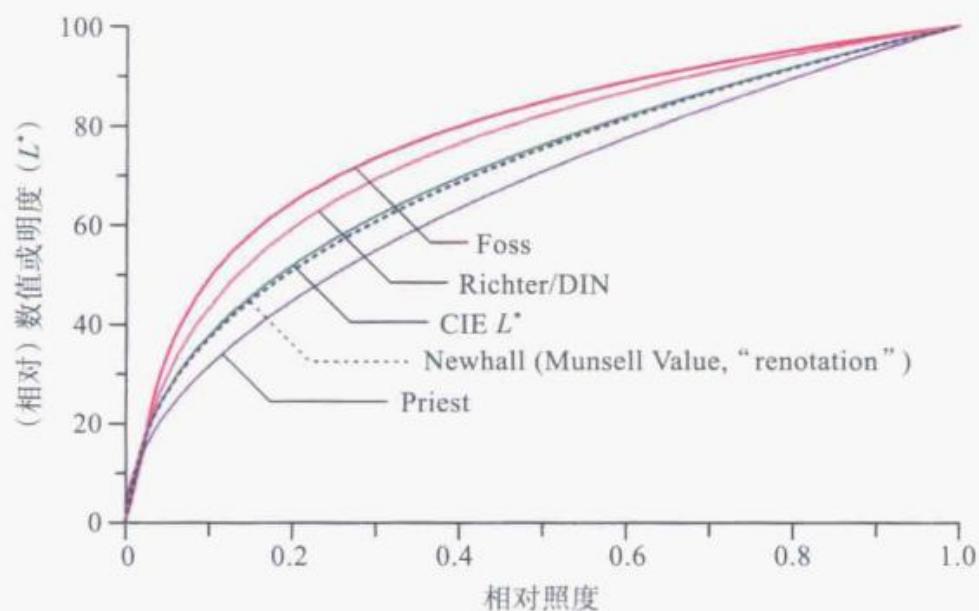


图 24-2 照度和明度

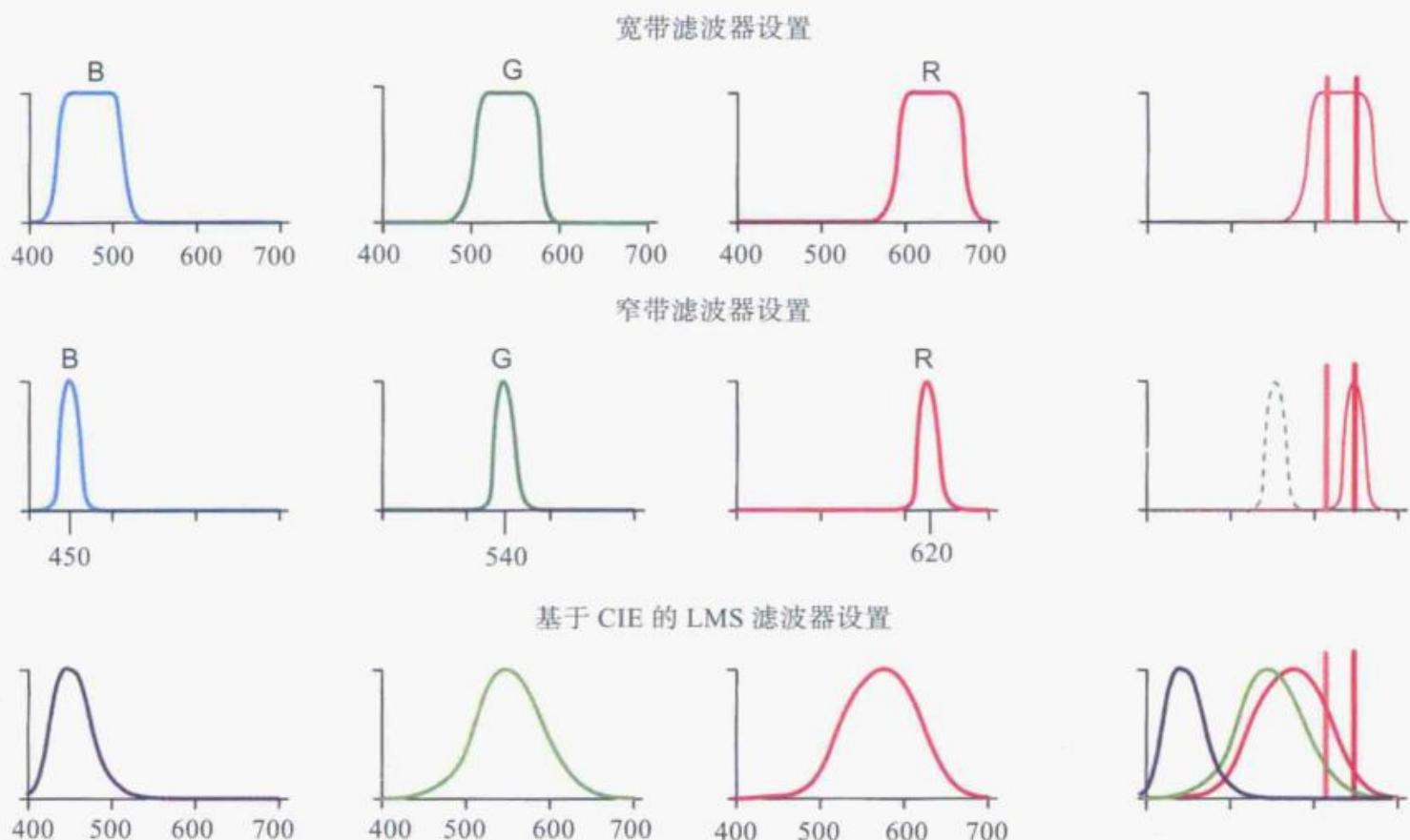


图 25-3 光谱约束图

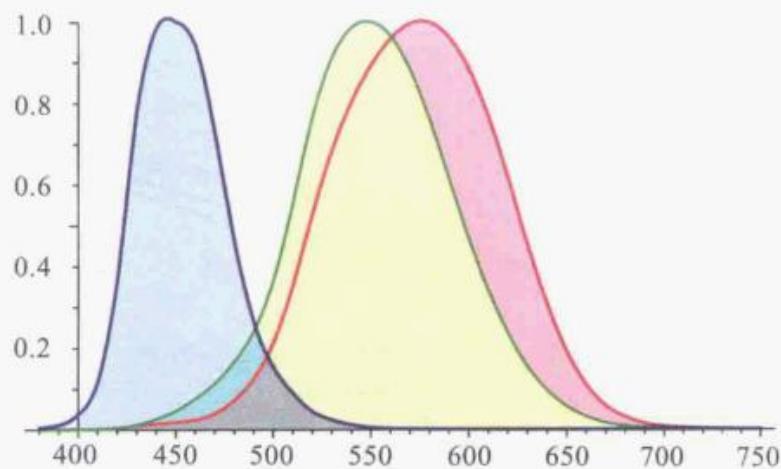


图 25-4 HPE 配色函数

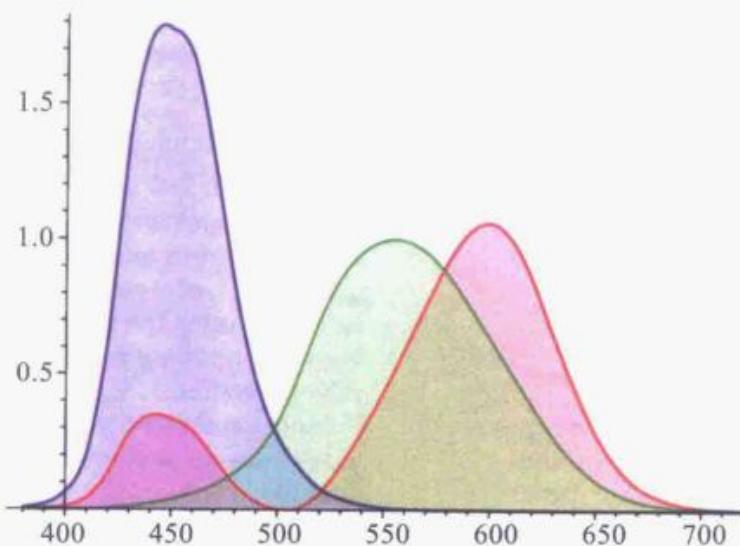


图 25-5 CIE 1931 2° 配色函数

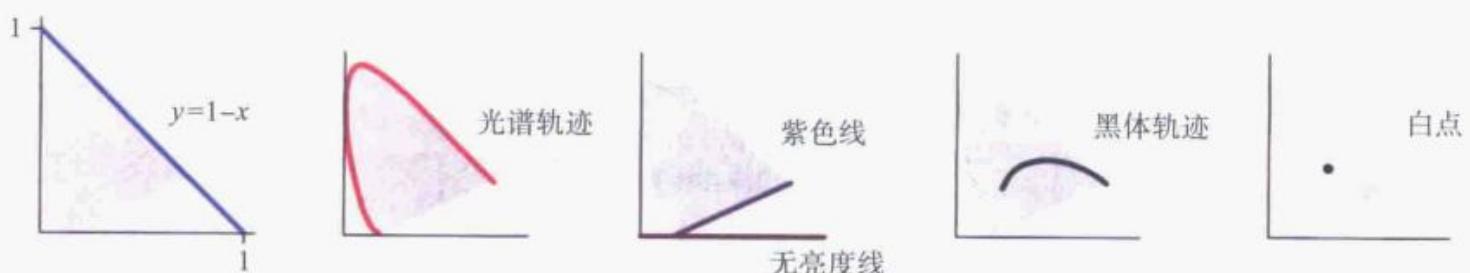


图 25-8 CIE $[x, y]$ 特征图

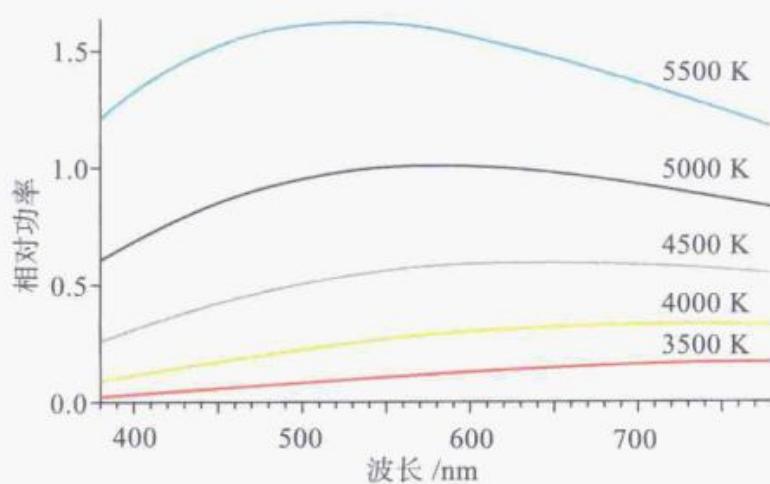


图 25-9 黑体辐射在几个温度点上的 SPD 绘于图中

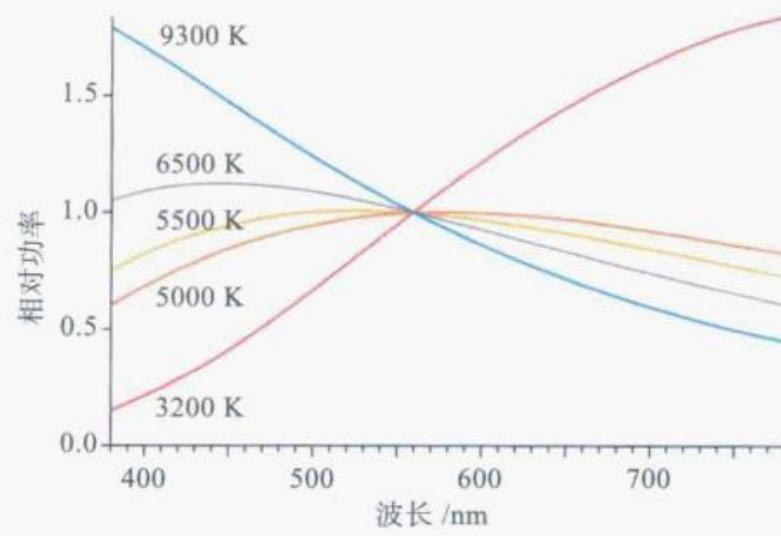


图 25-10 规范化黑体辐射 SPD 图中

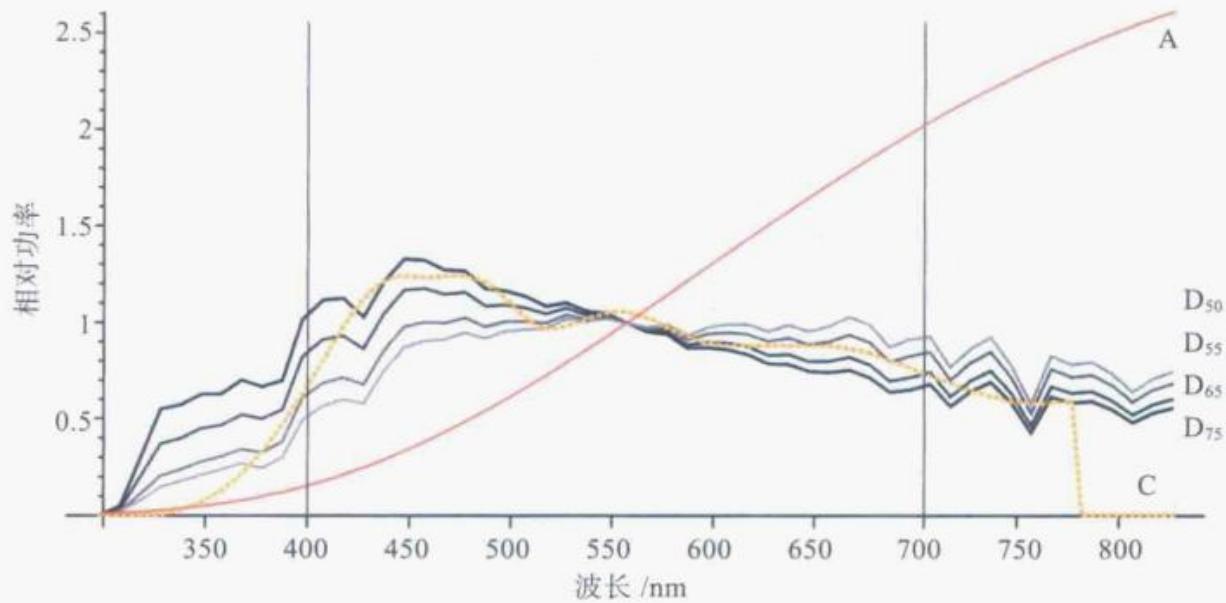


图 25-11 CIE 辐射体

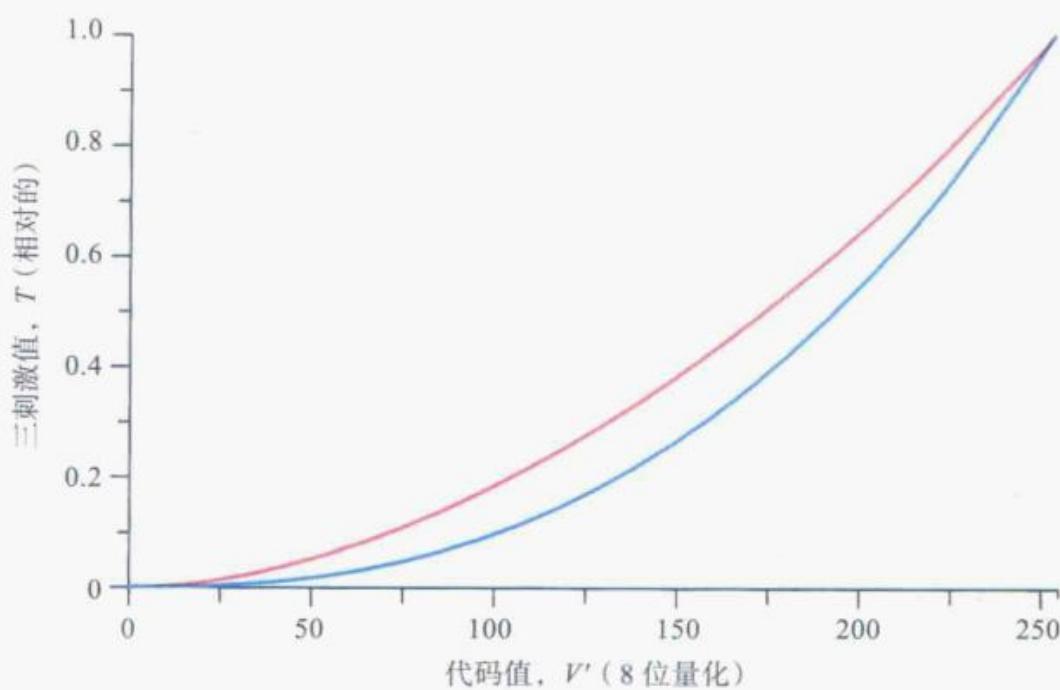


图 27-6 通用计算机与经典苹果计算机的伽马值

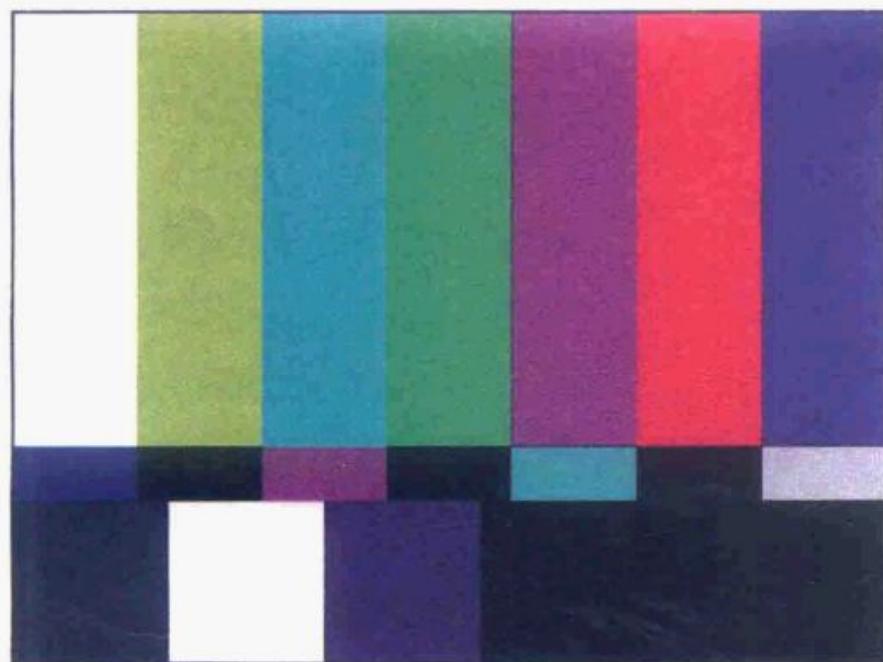


图 36-1 SMPTE EG 1 标清彩条测试信号

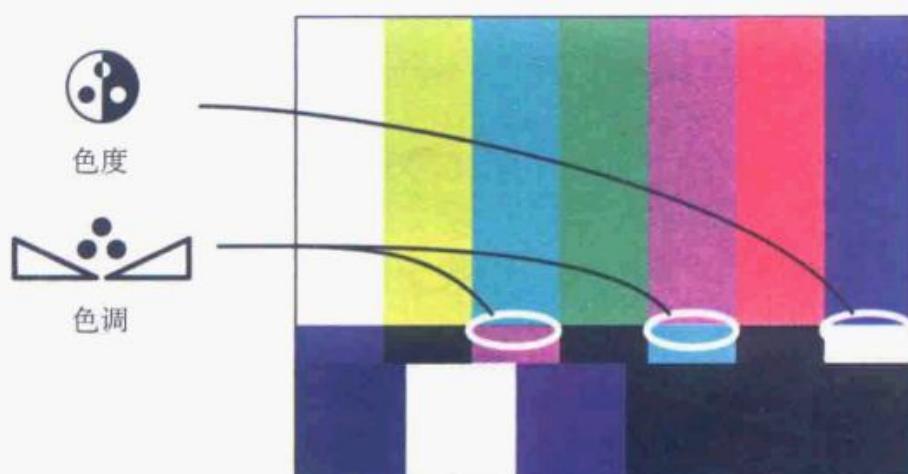


图 36-4 调整色调和色度

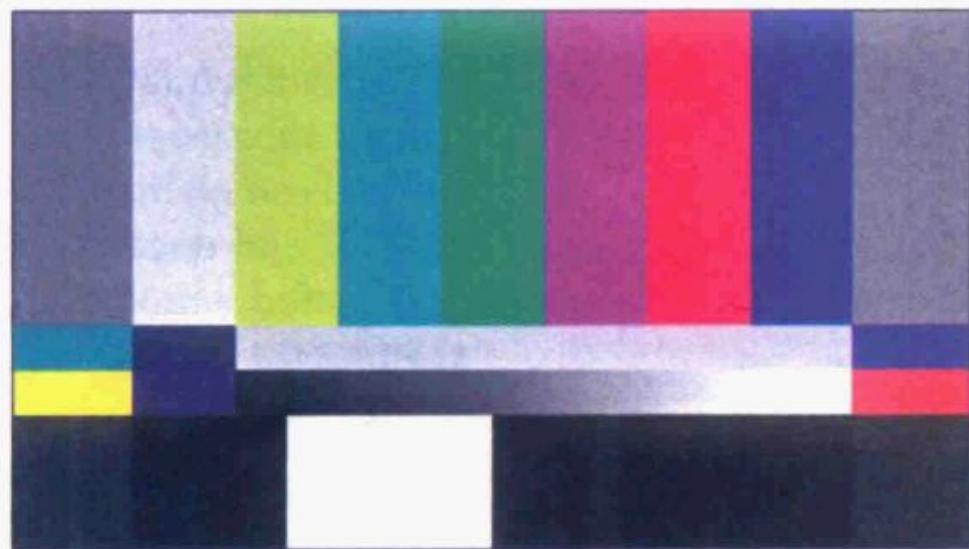


图 36-5 SMPTE RP 219 高清彩条测试信号

表 18-1 色彩基色

代 码	解 释
0	已禁用
1	BT.709
2	未指定
3	已保留 / 用于将来
4	BT-470-6/NTSC 1953
5	EBU Tech. 3213
6	SMPTE EP 145
7	SMPTE 240M
8	"Generic film"

表 18-2 传输特性

代 码	解 释
0	已禁用
1	BT.709
2	未指定
3	已保留 / 用于将来
4	显示伽马值 2.2
5	显示伽马值 2.8
6	BT.709
7	SMPTE 240M
8	Linear
9	Log (10 ² : 1)
10	Log (10 ^{2.5} : 1)
11	xvYCC
12	BT.1361
13, …, 255	已保留

表 18-3 矩阵系数

代 码	解 释
0	已禁用
1	BT.709
2	未指定
3	已保留 / 用于将来
4	BT.601
5	BT.601
6	BT.601
7	SMPTE 240M
8	$Y'C_G C_O$
9, …, 255	已保留

表 18-4 色彩基色的解释

代 码	解 释
0	已禁用
1	用于 HD 或 SD
2	未指定
3	已保留 / 用于将来
4	使用代码 1
5	用于 SD
6	用于 SD
7	使用代码 1
8	使用代码 1

表 18-5 传输特性的解释

代 码	解 释
0	已禁用
1	使用
2	未指定
3	已保留 / 用于将来
4	使用代码 1
5	使用代码 1
6	用代码 1 代替
7	使用代码 1
8	使用代码 1
9	使用代码 1
10	使用代码 1
11	xvYCC: 实现
12	使用代码 1
13, …, 255	已保留

注：代码 11 xvYCC 有可能会成为未来的系统，但是现在不太可能实现。

表 18-6 矩阵系数解释

代 码	解 释
0	已禁用
1	用于 HD
2	未指定
3	已保留 / 用于将来
4	用于 SD
5	用于 SD
6	用于 SD
7	使用代码 1
8	$Y'C_GC_0$: 实现
9, …, 255	已保留

注：代码 8 ($Y'C_GC_0$) 可能会用于未来的系统。我不知道现在的解码器能不能解码 $Y'C_GC_0$ 的色彩空间，不过明显没有好处。现在来说也极不可能实现。

出版者的话

文艺复兴以来，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，信息学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅擘划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的信息产业发展迅猛，对专业人才的需求日益迫切。这对我国教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其信息科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀教材将对我国教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章公司较早意识到“出版要为教育服务”。自 1998 年开始，我们就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与 Pearson、McGraw-Hill、Elsevier、John Wiley & Sons、CRC、Springer 等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出 Alan V. Oppenheim Thomas L. Floyd、Charles K. Alexander、Behzad Razavi、John G. Proakis、Stephen Brown、Allan R. Hambley、Albert Malvino、Peter Wilson、H. Vincent Poor、Hassan K. Khalil、Gene F. Franklin、Rex Miller 等大师名家的经典教材，以“国外电子与电气技术丛书”和“国外工业控制与智能制造丛书”为系列出版，供读者学习、研究及珍藏。这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也越来越多被实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着电气与电子信息学科建设的不断完善和教材改革的逐渐深化，教育界对国外电气与电子信息教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章公司欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方法如下：

华章网站：www.hzbook.com

电子邮件：hzjsj@hzbook.com

联系电话：(010) 88379604

联系地址：北京市西城区百万庄南街 1 号

邮政编码：100037



华章教育

华章科技图书出版中心

译者序

图像是人类认识世界并且传递信息的重要技术手段，数字视频就是以数字形式记录的连续图像信息。随着计算机技术和智能设备的快速发展，数字视频与高清技术已经渗透到人们的日常生活、工作和社会生产中。数字视频技术的发展必然会影响人们生活和工作的方方面面，随着科学技术的不断发展，数字视频技术的应用领域将会不断扩大。

本书从基本概念、基础理论、实际问题、演播室标准、视频压缩、发行标准 6 个大的方面系统阐述了数字视频与高清技术。作者多年从事数字视频与高清技术方面的学术及工程应用研究，在书中对基本概念的诠释极为精准，这使得本书在同类书籍中脱颖而出。希望本书中文版的发行能吸引我国更多的科研人员从事相关的理论研究和系统研制，提高国内学者在数字视频和高清技术方面的基础素养，为将来的创新研究打下基础。

全书由褚晶辉、吕卫、马永涛和宫霄霖共同翻译，全书的译文由刘开华教授统一术语、格式，并进行一致性检查和统稿。在翻译的过程中，译者对原著关于图的细微描述做了删减，对原著的排版格式进行了修改，以适应国内读者的习惯。刘开华教授课题组的研究生薛蒙蒙、付伟、汪忠伟、徐文会、张忍、王晶、沈郭浩、申倩伟、曾亚辉对本书翻译做出很大贡献，在此表示谢意！感谢机械工业出版社积极引进本书，使得本书中文版能够尽快与读者见面。

在本书的翻译过程中，我们力求忠实、准确地表达原著的本意，同时保留原著的写作风格。尽管数字视频与高清技术是译者从事科学研究的重要方向，但由于水平有限，译文中难免存在一些错误和不当之处，敬请广大读者批评指正。

本书适合作为高等院校计算机、电子与信息工程、通信工程、数字媒体等相关专业高年级本科生和研究生的“数字视频技术”课程教材或教学参考书；也可作为工程技术人员的参考资料或培训教材；还可供在视频技术、图像处理与通信、电视工程技术、影视技术、多媒体技术、电影技术等领域从事研究、开发、生产及维护的科技人员阅读、参考。

译 者

2017 年 4 月

前 言

自 2003 年年初本书第 1 版问世以来，视频技术不断发展，视频和计算机图形学两者进一步地“融合”（Jim Blinn 称之为碰撞）。电视正在没落，而计算机技术和网络传输正高奏凯歌。如今，哪怕“TV”这个缩写也是值得怀疑的：在过去的半个世纪中，TV 意味着广播，但今天大部分的视频——来自苹果的 iTunes 商店、Hulu、Netflix、YouTube 等——都不是传统意义上的广播。在这一版中，我用 SD（标清）和 HD（高清）取代 SDTV（标清电视）和 HDTV（高清电视）。

现在，数字视频无处不在，而图 P-1 所示的模拟扫描已过时。在这一版中，我把像素阵列提高到首要地位。第 1 版介绍了扫描线，而这一版改用图像行和图像列。这一版删除了微秒，改用样本计数；删除了毫伏，改用像素值。第 1 版中“巨大的数据容量”等已经被“相当大”或甚至“适中的数据容量”代替。

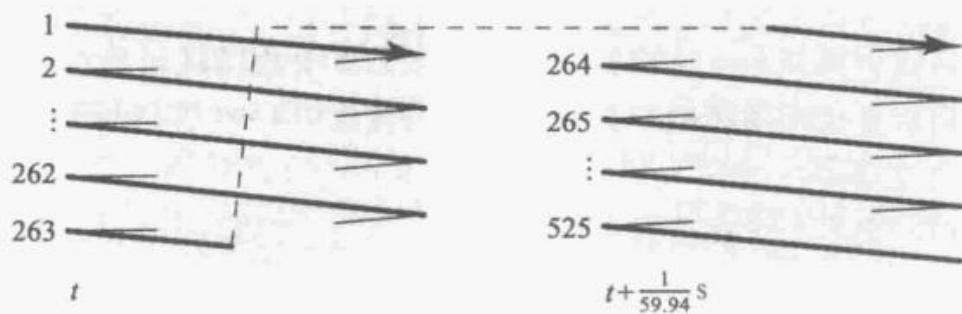


图 P-1 模拟扫描[⊖]

在我 1996 年出版的第一本书《数字视频技术介绍》和本书第 1 版中，我先介绍编码，然后介绍解码。从工程角度来看，这样的顺序是有道理的。然而，现在我认为这样的顺序里隐藏着一条深深的哲学裂痕。一旦节目素材准备就绪，经过解码并在演播室的参考显示器上播放，同时进行加工，也就是在下行流最终审定播出后，只有解码和显示过程才重要。为了在再次观看时显示一幅逼真的图像而进行图像数据采样和编码还算方便，但在加工时解码并显示图像数据的难度太高。如果创作人员在编码时为了达到某个美学效果而歪曲了颜色，例如，他们将黑电平提升 0.15 倍，或将色调旋转 123°——经典编码方程便不再适用，但这些图像数据的修改并不是编码出错引起的，而是创作意图运用的结果。传统的观念认为编码过程是固定的，但真正重要的是解码过程才是固定的。我提倡的原则很像 MPEG 的原则，精确定义了解码器，而只要产生的是合法的位流，编码器被允许做任何事。这一版里我强调解码。第 2 章作为新的一章，概述了这一哲学观点。

视频技术涉及的领域很广。对于本书重点阐述的图像编码，尤其是视觉科学、色彩学、图像学、信号处理与视频技术之间的关系。

如果别处有关于某一章主题的延伸、权威的信息，那么该章的结尾处会有“延伸阅

[⊖] 已经过时。在现代视频和高清中，图像为像素阵列的形式。任何著作若用此类图描述图像的采集或显示，都没有准确地描述数字视频。

读”部分。

SMPTE 是许多老式标准和现代标准的根源。在本书第 1 版和第 2 版相隔出版的时间里，SMPTE 弃用了许多历史标准的 M 后缀，而给标准加上 ST 前缀（并列使用 EG 表示工程指南，以及 RP 表示实践建议）。本书根据新的命名引用近期的 SMPTE 标准。

来到 2012 年，我们可以肯定地说，模拟电视技术和复合电视技术已经过时。写作本书第 1 版时，我把精力集中在我认为不会迅速改变的事情上。然而，第 1 版中可能有 15% 或 20% 的内容介绍了现在会被归类为“老式”的技术，主要是模拟电视和复合电视，以及 NTSC 和 PAL。由于对自己花了数百或数千小时写就的内容难以割舍，我虽然把这部分内容拿掉了，但却将其放在《Composite NTSC and PAL: Legacy Video Systems》书中，在 www.poynton.com/CNPLVS/ 上可免费获得。

据说一本书中每出现一个公式就会让潜在的读者减少一半。对于本书，我希望读者在计算过 10 个公式之后，数量下降到 2^{-10} ！本书保留了部分公式，但这些公式一般对于理解概念并非必需。如果你被某个公式吓到了，可跳过它，如果你愿意计算的话，也可以稍后再回来。我希望读者能用 Bringhurst 建议的关于页面组成原则的数学描述的方法来看待数学公式。在其经典著作《字体风格的要素》的第 8 章，Bringhurst 说道：“数学在此不是强加给任何人的苦差事。相反，它们在此完全是为了娱乐。它们的目的是为了愉悦那些也许是希望把事情做得更好而乐意检验自己正在做的或要做的或已做的事情的读者。那些在任何时候都更喜欢直接行动而把分析留给别人的读者，在本章中也会满足于研究图画而忽略文字。”

在写作本书时，我小心翼翼地避免引入错误。不过，尽管我和我的审稿人尽了最大的努力，但仍可能会有一些错误。就像我的前一本书一样，我将为这本书编写勘误表，并放在 www.poynton.com/DVAI2/ 中。请在此页面报告您发现的任何错误，我将努力订正它并把这归功于你！

Charles Poynton
2012 年 1 月于多伦多

致 谢

我对数字视频的入门是在很多年前为由 John Lowry 构想而由 Richard Kupnicki 设计的硬件写微代码。Ross Video 的创始人 Ross 继续了我在视频方面的教育。感谢这三个人。

应 CJOH 的工程副总裁 Austin Reeve 之邀，我花了很多时间在加拿大渥太华的 CJOH-TV 测试我的理论。感谢他的自信、幽默和耐心。

非常感激这几十年来不仅是我同事，更是我密友的 4 个人：C. R. Caillouet、Pierre Deguire、Charlie Pantuso 和 Mark Schubin。

感谢鼓励我参加这个项目的同事，他们都在不同阶段审阅了我的手稿：David Bancroft、George Joblove、Peter Symes，还有 Dave Lehoty——在很久以前就建议了本书所用的书名！

以下诸位审阅了部分手稿并指出了错误：Don Craig、Joseph Goldstone 和 Adam Wilt。也向他们致歉，因为这次无法把他们所有的建议都包含进来。

感谢为第 1 版提供错误报告的网友，特别是 Ken Greenebaum、Dragan Matković 和 Andrew Murray；特别感谢 Jay Zipnick 关于颜色计算的几点建议。

写作本书之时，我正在 Brian Funt 的指导下攻读博士学位。感谢他对本书的贡献，其中许多相关内容都可在本书中找到。

Diane Cerra 是我第一本书以及本书第 1 版的耐心细致的编辑。她的意见和辛勤工作使第 2 版成形。Sara Binns、Laura Lewin、Sara Scott 和 Graham Smith 是 Morgan Kaufmann 出版社负责本书的全体专业编辑，感谢他们。也感谢优秀的文字编辑 Charles Roumeliotis。

感谢我的家人，感谢他们的爱和鼓励：Peg、Al、Kim、Brenna、Alana、爸爸，还有 Jean。也感谢 Jeri、Corot、Ben 和 Paige。

感谢 Barbara，我生命中的爱，以及我们的女儿 Quinn 和 Georgia。

目 录

出版者的话
译者序
前言
致谢

第一部分 引言

第1章 光栅图像	2
1.1 幅型比概念	3
1.2 几何	4
1.3 图像采样	4
1.4 数字化	4
1.4.1 量化	5
1.4.2 一维采样	5
1.4.3 二维采样	5
1.5 感知均匀性	5
1.6 色彩	6
1.7 亮度和色差分量	6
1.8 数字图像的表示方法	7
1.9 标清和高清	8
1.10 方形采样	8
1.11 幅型比的比较	9
1.12 幅型比	9
1.13 帧频	11
第2章 图像的采样和显示	12
2.1 图像状态	12
2.2 EOCF 标准	13
2.3 娱乐节目	14
2.4 图像获取	14
2.5 面向消费者的视频获取	14
2.6 消费类电子显示器	15
第3章 线性光和感知均匀性	16
3.1 对比	16
3.2 对比度	17
3.3 感知均匀性	18
3.4 “编码 100” 难题和非线性	
图像编码	18

3.5 线性和非线性	21
第4章 量化	22
4.1 线性度	22
4.2 分贝	22
4.3 噪声、信号和灵敏度	23
4.4 量化误差	24
4.5 全摆幅	24
4.6 演播室摆幅（下余量和上余量）	25
4.7 接口偏移量	26
4.8 处理编码	26
4.9 二进制补码回绕	27
第5章 对比度和明亮度	28
5.1 感知特性	28
5.2 显示处理信号的历史	28
5.3 数字驱动电平	29
5.4 信号和明度之间的关系	30
5.5 算法	31
5.6 黑色电平设置	33
5.7 对比度和明亮度控制的影响	33
5.8 交替传译	35
5.9 LCD 中的明亮度和对比度	
控制	36
5.10 PDP 中的明亮度和对比度	
控制	36
5.11 桌面图形的明亮度和对比度	
控制	36
第6章 计算中的光栅图像	38
6.1 特征图像的描述	39
6.2 光栅图像	39
6.2.1 二值图像	39
6.2.2 灰度图像	40
6.2.3 真彩色图像	40
6.2.4 伪彩色图像	41
6.3 不同类型之间的转换	41
6.4 图像文件	42
6.5 计算机图形学中的“分辨率”	42

第 7 章 图像结构	43	12.2 照度简述	69
7.1 图像重建	43	12.3 色差编码(色度)	69
7.2 采样孔径	44	12.3.1 $Y'P_R P_B$	70
7.3 光斑轮廓	45	12.3.2 $Y'C_R C_B$	70
7.4 “箱”分布	46	12.3.3 $Y'UV$	70
7.5 高斯分布	46	12.3.4 $Y'IQ$	70
第 8 章 光栅扫描	47	12.4 色度亚采样	71
8.1 闪烁率、刷新率和帧频	47	12.4.1 4:4:4	71
8.2 扫描简介	48	12.4.2 4:2:2	71
8.3 扫描参数	48	12.4.3 4:1:1	71
8.4 隔行扫描	49	12.4.4 4:2:0	71
8.5 抖动	50	12.5 色度亚采样符号	71
8.6 模拟系统中的隔行扫描	51	12.6 色度亚采样滤波器	72
8.7 隔行扫描和逐行扫描	51	12.7 复合 NTSC 和 PAL 的色度	73
8.8 扫描表示法	52	第 13 章 SD 分量介绍	74
8.9 运动描绘	53	13.1 扫描标准	74
8.10 分段帧(24PsF)	53	13.2 宽屏(16:9)SD	76
8.11 视频系统的分类	53	13.3 方形和非方形采样	76
8.12 系统之间的转换	53	13.4 重采样	77
第 9 章 分辨率	55	第 14 章 复合 NTSC 和 PAL 的介绍	78
9.1 幅频响应和带宽	55	14.1 NTSC 和 PAL 编码	78
9.2 视觉灵敏度	56	14.2 NTSC 和 PAL 解码	79
9.3 观看距离和角度	56	14.3 S 视频接口	80
9.4 克尔效应	57	14.4 频率交织	80
9.5 分辨率	58	14.5 复合模拟 SD	80
9.6 视频中的分辨率	59	第 15 章 HD 介绍	81
9.7 观看距离	59	15.1 HD 扫描	81
9.8 再谈隔行扫描	59	15.2 BT.709 HD 彩色编码	83
第 10 章 恒定照度	61	第 16 章 视频压缩简介	84
10.1 恒定照度原理	61	16.1 数据压缩	84
10.2 CRT 补偿	62	16.2 图像压缩	84
10.3 从恒定照度的分离	62	16.3 有损压缩	84
10.4 亮度	63	16.4 JPEG	85
10.5 照度转换成色度的“泄露”	63	16.5 运动 JPEG	85
第 11 章 图片渲染	66	16.6 JPEG 2000	86
11.1 环绕效果	66	16.7 夹层压缩	86
11.2 色调比例调整	66	16.8 MPEG	87
11.3 渲染合并	67	16.9 图像编码类型(I、P、B)	87
11.4 桌面计算中的图像渲染	68	16.10 重排	89
第 12 章 亮度和色度简介	69	16.11 MPEG-1	89
12.1 亮度	69	16.12 MPEG-2	89

16.13	其他 MPEG 标准	90
16.14	MPEG IMX	90
16.15	MPEG-4	90
16.16	H.264	90
16.17	帧内 AVC	91
16.18	WM9、WM10、VC-1 编解码	91
16.19	CE 采样压缩	91
16.20	HDV	91
16.21	AVCHD	92
16.22	IP 传输到用户的压缩	92
16.23	VP8 (WebM) 编解码	92
16.24	狄拉克 (基本)	92
第 17 章 流和文件		93
17.1	历史概述	93
17.2	物理层	94
17.3	流接口	94
17.4	IEEE 1394 (火线接口、 i.LINK)	95
17.5	HTTP 实时流媒体(HLS)	96
第 18 章 元数据		97
18.1	元数据例 1: CD-DA	97
18.2	元数据例 2: .yuv 文件	98
18.3	元数据例 3: RFF	98
18.4	元数据例 4: JPEG 或 JFIF	99
18.5	元数据例 5: 序列显示扩展	100
18.6	小结	102
第 19 章 立体 (三维) 视频		104
19.1	图像获取	104
19.2	S3D 显示	104
19.3	立体电影	104
19.4	时间多路复用	105
19.5	偏光法	105
19.6	波长复用 (光谱立体/杜比)	106
19.7	裸眼立体显示	106
19.8	视差屏障显示	106
19.9	立体光栅显示	106
19.10	录制和压缩	107
19.11	用户接口和显示	107
19.12	幻影	108
19.13	聚散度和调节	108

第二部分 理论

第 20 章 滤波和采样		110
20.1	采样定理	110
20.2	对频率为 $0.5f_s$ 的信号进行 采样	111
20.3	幅频响应	112
20.4	矩形函数的幅频响应	113
20.5	sinc(·) 加权函数	114
20.6	点采样的频率响应	114
20.7	傅里叶变换对	115
20.8	模拟滤波器	116
20.9	数字滤波器	116
20.10	冲激响应	119
20.11	有限冲激响应滤波器	119
20.12	滤波器的实际可实现性	120
20.13	相位响应 (群延迟)	121
20.14	无限冲激响应滤波器	121
20.15	低通滤波器	122
20.16	数字滤波器设计	124
20.17	信号重构	125
20.18	接近 $0.5f_s$ 的重构	125
20.19	" $(\sin x)/x$ " 修正	126
延伸阅读		127
第 21 章 重采样、插值和抽取		128
21.1	2:1 下采样	129
21.2	过采样	130
21.3	插值	131
21.4	拉格朗日插值	131
21.5	用于滤波的拉格朗日插值	132
21.6	多相插值器	133
21.7	多相抽头和相位	134
21.8	多相插值器的实现	134
21.9	抽取	135
21.10	用于抽取的低通滤波器	135
第 22 章 图像数字化与重构		136
22.1	空间频域	137
22.2	梳状滤波	138
22.3	空域滤波	139
22.4	图像预采样滤波器	139
22.5	图像重构滤波器	139

22.6 空间(二维)过采样	140	26.11 照度系数	175
第23章 感知和视敏特性	142	26.12 RGB 和 CIE XYZ 之间的 转换	176
23.1 视网膜	142	26.13 矩阵化引入的噪声	177
23.2 适应性	142	26.14 RGB 系统之间的转换	177
23.3 对比灵敏度	143	26.15 相机白色基准	178
23.4 对比灵敏度函数	144	26.16 显示器白光基准	178
第24章 照度和明度	146	26.17 色域	178
24.1 辐射度和强度	146	26.18 宽色域重现	179
24.2 照度	147	26.19 自由尺度色域、自由尺度 指数	179
24.3 相对照度	148	延伸阅读	179
24.4 红、绿、蓝的照度	148		
24.5 亮度(CIE L^*)	148		
第25章 CIE 色度系统	151	第27章 伽马值	181
25.1 视觉的基本原理	151	27.1 CRT 物理学中的伽马值	181
25.2 定义	151	27.2 惊人的巧合	182
25.3 光谱功率分布和三刺激值	152	27.3 视频系统中的伽马校正	182
25.4 光谱约束	153	27.4 光电转换函数	183
25.5 CIE XYZ 三刺激值	155	27.5 BT.709 OECF	184
25.6 CIE $[x, y]$ 色品	156	27.6 SMPTE 240M OECF	185
25.7 黑体辐射	157	27.7 sRGB 传递函数	185
25.8 色温	158	27.8 标清的传递函数	186
25.9 白光	158	27.9 位深度的需求	187
25.10 色彩适应	159	27.10 现代显示设备中的伽马 校正	187
25.11 感知均匀颜色空间	160	27.11 伽马估计	188
25.12 CIE $L^* u^* v^*$ 系统	160	27.12 视频、CGI 和 Macintosh 中的 伽马值	189
25.13 CIE $L^* a^* b^*$ 参数 (CIE LAB)	161	27.13 计算机图形学中的伽马值	191
25.14 CIE $L^* u^* v^*$ 和 CIE $L^* a^* b^*$ 总结	162	27.14 伪彩色中的伽马值	192
25.15 色彩规范和彩色图像编码	162	27.15 8 位线性编码的局限	192
延伸阅读	163	27.16 CGI 中的线性和非线性 编码	192
第26章 视频色彩学	164	第28章 亮度与色差	194
26.1 加性重建(RGB)	164	28.1 色敏度	194
26.2 RGB 基色特性	165	28.2 RGB 和 $R'G'B'$ 彩色立方体	194
26.3 BT.709 基色	166	28.3 传统的亮度、色差编码	197
26.4 老式标清基色	167	28.4 照度和亮度标记	198
26.5 sRGB 系统	168	28.5 非线性红、绿、蓝 ($R'G'B'$)	199
26.6 SMPTE 无标度基色	168	28.6 BT.601 亮度	200
26.7 AMPAS ACES 基色	168	28.7 BT.709 亮度	200
26.8 SMPTE/DCI P3 基色	169	28.8 再论色度亚采样	200
26.9 CMF 和 SPD	169		
26.10 归一化和缩放	175		

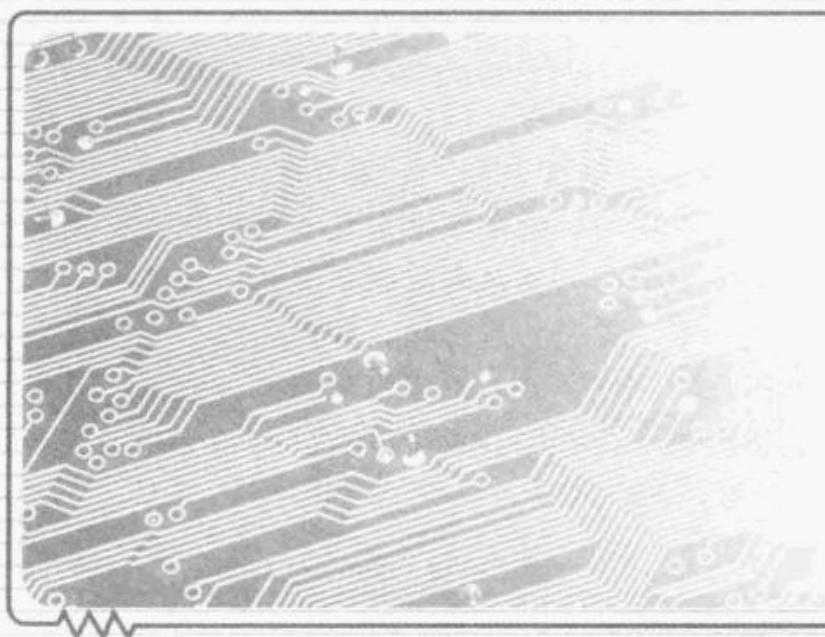
28.9 亮度/色差总结	200	32.2 行频	225
28.10 标清和高清亮度的混乱	202	32.3 声音副载波	225
28.11 亮度、色差分量集	203	32.4 添加复合彩色	226
第三部分 实际问题			
第 29 章 标清分量视频彩色		第 30 章 高清分量视频彩色	
编码	206	编码	213
29.1 标清的 $B'-Y'$ 、 $R'-Y'$ 分量	207	30.1 BT. 709 高清 $B'-Y'$ 和 $R'-Y'$ 分量	213
29.2 标清的 $P_B P_R$ 分量	207	30.2 BT. 709 高清 $P_B P_R$ 分量	214
29.3 标清的 $C_B C_R$ 分量	208	30.3 BT. 709 高清 $C_B C_R$ 分量	214
29.4 从演播室 RGB 到 $Y' C_B C_R$	210	30.4 xvYCC 的 $C_B C_R$ 分量	215
29.5 从计算机 RGB 到 $Y' C_B C_R$	210	30.5 从演播室 RGB 到 $Y' C_B C_R$	215
29.6 “全摆幅” $Y' C_B C_R$	210	30.6 从计算机 RGB 到 $Y' C_B C_R$	216
29.7 $Y'UV$ 、 $Y'IQ$ 的混淆	211	30.7 高清和标清之间的转换	216
第 31 章 视频信号处理	218	30.8 彩色编码标准	217
31.1 边沿处理	218		
31.2 过渡样值	218		
31.3 图像行	219		
31.4 SAL 和 SPW 参数选择	219		
31.5 视频电平	220		
31.6 建立电平（基座）	220		
31.7 从 BT. 601 到计算机领域	221		
31.8 增强	221		
31.9 中值滤波	222		
31.10 核心化	223		
31.11 色度转换改善	223		
31.12 混合与键控	223		
第 32 章 帧频、场频、行频和采样率	225		
32.1 场频	225		
32.2 行频	225		
32.3 声音副载波	225		
32.4 添加复合彩色	226		
32.5 NTSC 彩色副载波	226		
32.6 576i PAL 彩色副载波	227		
32.7 $4f_{sc}$ 采样	227		
32.8 常见的采样率	227		
32.9 高清扫描的数字学	228		
32.10 音频速率	229		
第 33 章 时间码	230		
33.1 引言	230		
33.2 丢弃帧时间码	230		
33.3 编辑	231		
33.4 线性时间码	231		
33.5 垂直间隔时间码	232		
33.6 时间码结构	232		
延伸阅读	232		
第 34 章 2-3 下拉	234		
34.1 2-3-3-2 下拉	235		
34.2 将电影转换到不同的帧频	235		
34.3 原生 24Hz 编码	237		
34.4 转换到其他帧频	237		
第 35 章 去隔行	239		
35.1 空间域	239		
35.2 垂直-时间域	240		
35.3 运动自适应性	241		
延伸阅读	242		
第 36 章 彩条	243		
36.1 标清彩条	243		
36.2 标清彩条标记法	244		
36.3 PLUGE 元素	244		
36.4 使用彩条调整复合解码器	244		
36.5 标清彩条中的 $-I$ 、 $+Q$ 和 PLUGE 元素	245		
36.6 高清彩条	245		
第四部分 演播室标准			
第 37 章 参考显示器和观看条件	248		
37.1 简介	248		
37.2 信号接口	248		
37.3 参考基色（黑色、白色）	248		

37.4 参考 EOCF	248	40.3 行同步	268
37.5 参考观看条件	249	40.4 模拟场/帧同步	268
延伸阅读	249	40.5 $R'G'B'$ EOCF 和三基色	270
第 38 章 SDI 和 HD-SDI 接口	250	40.6 亮度 (Y')	270
38.1 分量数字标清接口 (BT. 601)	250	40.7 图像中心、幅型比和消隐	270
38.2 串行数字接口	252	40.8 数字分量 4 : 2 : 2 接口	270
38.3 分量数字 HD-SDI	252	40.9 模拟分量 576i 接口	271
38.4 SDI 和 HD-SDI 同步、TRS 和 辅助数据	253	第 41 章 1280 × 720 像素高清	272
38.5 4 : 2 : 2 SD-SDI 中的 TRS	253	41.1 扫描	272
38.6 HD-SDI 中的 TRS	255	41.2 模拟同步	272
38.7 模拟同步和数字/模拟时序 关系	255	41.3 图像中心、幅型比和消隐	273
38.8 辅助数据	255	41.4 $R'G'B'$ EOCF 和三基色	273
38.9 SDI 编码	256	41.5 亮度 (Y')	274
38.10 HD-SDI 编码	257	41.6 数字分量 4 : 2 : 2 接口	274
38.11 压缩视频接口	257	第 42 章 1920 × 1080 像素高清	275
38.12 SDTI	257	42.1 扫描	275
38.13 切换和混合	258	42.2 模拟同步	276
38.14 数字设备的时序	258	42.3 图像中心、幅型比和消隐	278
38.15 ASI	258	42.4 $R'G'B'$ EOCF 和三基色	278
38.16 数字接口总结	259	42.5 亮度 (Y')	279
第 39 章 480i 分量视频	260	42.6 数字分量 4 : 2 : 2 接口	279
39.1 帧频	260	第 43 章 高清录像带	280
39.2 隔行扫描	260	43.1 D-5 HD (HD-D5、D-15)	280
39.3 行同步	261	43.2 D-6	281
39.4 场/帧同步	262	43.3 HDCAM (D-11)	281
39.5 $R'G'B'$ EOCF 和三基色	263	43.4 DVCPRO HD (D-12)	281
39.6 亮度 (Y')	263	43.5 HDCAM SR (D-16)	281
39.7 图像中心、幅型比和消隐	263	第 44 章 模拟分量高清接口	282
39.8 半行消隐	264		
39.9 数字分量 4 : 2 : 2 接口	264	第五部分 视频压缩	
39.10 模拟分量 $R'G'B'$ 接口	265		
39.11 EBU N10 模拟分量 $Y'P_BP_R$ 接口	265	第 45 章 JPEG 和 M-JPEG 压缩	286
39.12 行业标准模拟分量 $Y'P_BP_R$ 接口	266	45.1 JPEG 块和 MCU	286
第 40 章 576i 分量视频	267	45.2 JPEG 框图	287
40.1 帧频	267	45.3 电平转换	288
40.2 隔行扫描	267	45.4 离散余弦变换	288
		45.5 JPEG 编码示例	288
		45.6 JPEG 解码	290
		45.7 压缩率控制	290
		45.8 JPEG/JFIF	291
		45.9 M-JPEG	292
		延伸阅读	292

第 46 章 DV 压缩	293	48.8 灵活的运动补偿	314
46.1 DV 色度亚采样	293	48.9 四分之一像素运动补偿 插值	314
46.2 DV 帧/场模式	294	48.10 运动补偿预测的加权和 偏置	315
46.3 DV 的穿梭图像	295	48.11 16 位整数变换	315
46.4 DV 溢出方案	295	48.12 量化器	315
46.5 DV 量化	295	48.13 可变长编码	315
46.6 DV 数字接口	296	48.14 上下文自适应性	316
46.7 消费类 DV 录制	297	48.15 CABAC	317
46.8 专业 DV 变体	297	48.16 去块滤波器	317
第 47 章 MPEG-2 视频压缩	298	48.17 缓冲区控制	317
47.1 MPEG-2 的类和级	298	48.18 可分级视频编码	317
47.2 图像结构	300	48.19 多视点视频编码	318
47.3 MPEG 中的帧频和 2-3 下拉 ...	301	48.20 AVC-Intra	318
47.4 亮度和色度采样结构	301	延伸阅读	318
47.5 宏块	302	第 49 章 VP8 压缩	319
47.6 图像编码类型 (I、P、B)	302	延伸阅读	320
47.7 预测	303		
47.8 运动向量	304		
47.9 块编码	305		
47.10 帧和场 DCT 类型	305		
47.11 之字形扫描和可变长编码 ...	306		
47.12 刷新	307		
47.13 运动估计	307		
47.14 码率控制和缓冲区管理	308		
47.15 位流句法	309		
47.16 传输	310		
延伸阅读	310		
第 48 章 H. 264 视频压缩	311		
48.1 算法特性、类和级	311		
48.2 基本类和扩展类	313		
48.3 高类	313		
48.4 层次结构	313		
48.5 多参考图像	314		
48.6 宏块条	314		
48.7 空间帧内预测	314		
		第六部分 分发标准	
		第 50 章 MPEG-2 存储与传输	322
		50.1 基本流	322
		50.2 打包的基本流	322
		50.3 MPEG-2 节目流	322
		50.4 MPEG-2 传送流	323
		50.5 系统时钟	323
		延伸阅读	324
		第 51 章 数字电视广播	325
		51.1 日本	325
		51.2 美国	325
		51.3 ATSC 调制	326
		51.4 欧洲	327
		延伸阅读	327
		附录 A 使用 YUV 和照度是有害的 ...	328
		附录 B 辐射测定与光度测定简介 ...	332

第一部分

引 言



第1章

光栅图像

一幅数字图像可以由图像元素（像素，或像素点）的矩形阵列（矩阵）表示。几个图像标准的像素阵列如图 1-1 所示。在一个灰度系统中，每个像素包含一个单一的组成部分，它的值称为亮度。在一个彩色系统中，每个像素由几个部分组成。通常情况下，每个像素由三个部分组成，每个部分的值都与人类色彩感知系统密切相关。

从历史上看，视频图像由相机获取，通过信道传输，并使用模拟扫描显示，没有明确的像素矩阵。现代的照相机和显示仪器直接反映了图像矩阵的离散元素，其具有固定的结构。信号处理器位于相机，信道或者显示仪器中，它可以执行空间上和（或者）时间上的重采样来适应不同的格式。

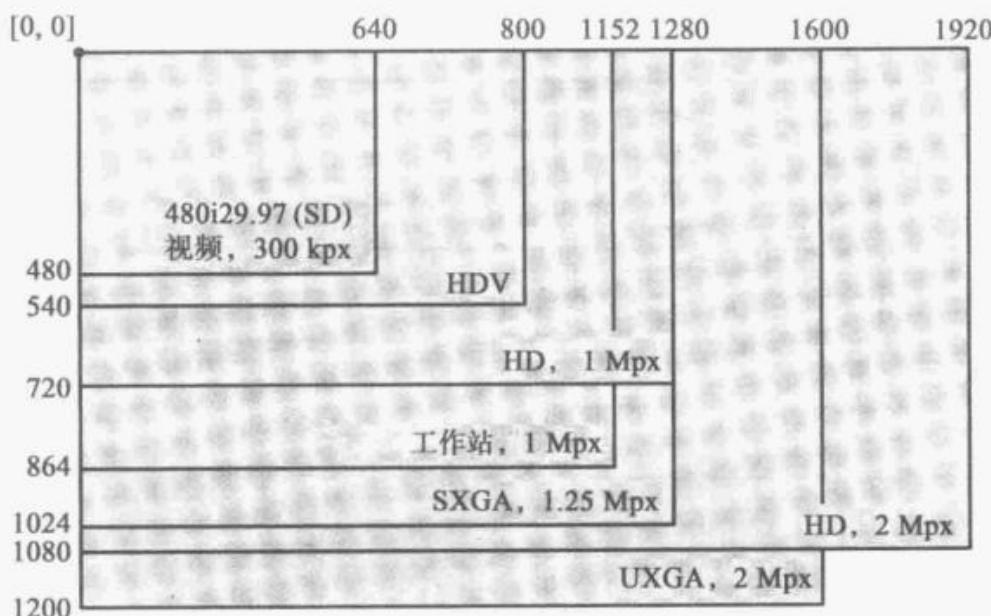


图 1-1 不同图像标准的像素阵列

一幅图像的像素矩阵是一帧。在视频中，用于存储一幅图像的数字存储器，称为一个帧存储器；在计算机中，这是一个帧缓冲区。一幅图像的总像素数目等于列像素数 N_C （或者是视频的有效行的采样率 S_{AL} ）乘以行像素数 N_R （或者是有效行 L_A ）。总的像素数目通常由百万像素（Mpx）表示。

在视频和计算中，一个像素包含代表颜色（通常为红色、绿色和蓝色）所必需的所有信息。在典型数码相机（DSC）的马赛克传感器中，一个像素是任意颜色的分量。去马赛克的过程就是插入缺失的成分，从而建立一个完整的图像阵列的过程。数字电影摄像机使用了数码相机（DSC）插入像素的方法，但是在数字电影放映机中，像素是一个三元组。

每个像素的分量值表示为采样栅格中对应点周围的一个小区域的亮度和色彩。

像素的分量值通常被量化成一个整数值，占用 1~16 位的数据存储空间（通常占用 8 位或 10 位）。每一个分量或者每一个像素的位数目称为位深度（为了避免混淆，我们使用位深度而不是位宽度。宽度是指整个图片）。

1.1 帧型比概念

简单地说，帧型比是指一副图像的宽度和高度的比例。电影和视频的标准的帧型比如图 1-2 所示。本书中所说的帧型比有时称为显示帧型比 (DAR) 或图片帧型比 (PAR)。标准清晰度 (SD) 电视的帧型比为 4 : 3。

视频图像	SD视频 4 : 3 1.33 : 1	宽屏SD视频 HD视频 16 : 9 ~1.78 : 1	
电影图像	35mm静止图像 3 : 2 1.5 : 1	影院电影 1.85 : 1	影院电影 2.4 : 1

图 1-2 视频、高清、电影的帧型比

$$\frac{\text{宽度}}{\text{高度}} = \frac{\text{AR}}{\text{SAR}} \quad (1-1)$$

$$N_C = \sqrt{n \text{AR}}; N_R = \sqrt{\frac{n}{\text{AR}}} \quad (1-2)$$

式 (1-1) 是关于图像和样本的帧型比。为了给帧型比为 AR 的图像赋值 n 阶方阵采样像素，根据式 (1-2)，选择图像列和图像的行进行计数（分别为 c 和 r ）。

电影胶片通常采用 1.85 : 1（由于历史原因，它被称为平面或者球面），或 2.4 : 1（“宽银幕电影”，或俗称宽银幕）。很多电影的帧型比是 1.85 : 1，但“大片”通常采用的是 2.4 : 1 帧型比。帧型比为 2.4 : 1 的电影在历史上是通过采用非球面透镜，以两倍因子压缩图像的水平维度获得的。投影机配备有一个类似的镜头来恢复投影图像的水平尺寸。这种透镜和技术称为变形。原则上，一个变形镜头可以有任意的比例，实际上，正好是 2 的比例普遍存在于电影应用中。

宽屏是指高于 4 : 3 的帧型比的屏幕。高清晰度 (HD) 电视标准帧型比为 16 : 9。在视频中，变形这个术语通常是指基本视频标准的 16 : 9 宽屏变形体，在这个变形体中 16 : 9 图像的水平尺寸占用了标准帧型比为 4 : 3 的相同宽度。消费电子设备很少恢复这样转换之后的正确帧型比（我们将在后面的章节探讨）

高清的标准化帧型比为 16 : 9 (约 1.78 : 1)，相当接近普通影片 1.85 : 1 的帧型比。图 1-3 显示的是帧型比为 16 : 9 的起源。通过一次数字命题的偶然机会 Kerns Powers 首次揭示 4 : 3 (传统电视的标准帧型比) 与 2.4 : 1 (宽银幕电影影片的帧型比) 的几何平均值非常接近于 16 : 9 (计算显示在图的右下角)。高清标准选择比例为 16 : 9 意味着标清、高清和宽银幕电影拥有共同的“影像圈”：16 : 9 是在

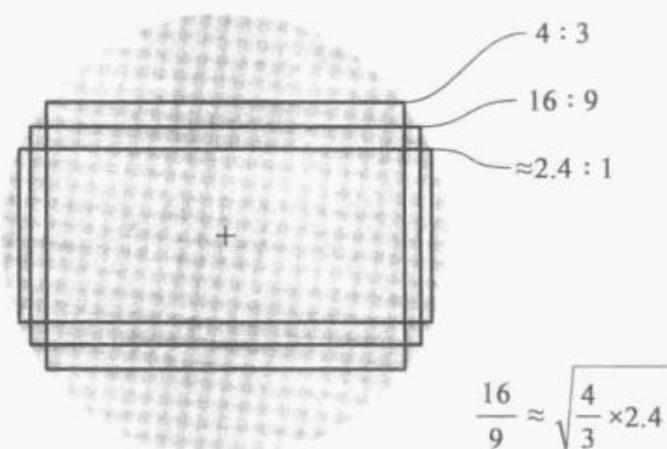


图 1-3 高清电视 16 : 9 的帧型比

高清电视选择 16 : 9 的帧型比是因为 16 : 9 非常接近传统的电视图像的 4 : 3 帧型比和宽银幕电影图像 2.4 : 1 的几何平均数

标清要求的垂直裁剪和宽银幕电影要求的水平裁剪中的折中选择。

1.2 几何

在数学中，平面（二维）的坐标值范围包括正值和负值。由此平面分为四个象限（见图 1-4）。象限在逆时针方向上用罗马数字表示。在连续图像平面中，使用直角坐标 (x, y) 表示位置——第一个坐标值代表水平方向，第二个坐标值代表垂直方向。当 x 和 y 都是正值时，位置位于第一象限（象限 I）。在图像学中，图像就在这个象限（Adobe 的 PostScript 系统使用第一象限坐标）。

在矩阵索引中，轴的顺序与笛卡儿坐标相反：先索引矩阵的行，再索引矩阵的列。矩阵的最上面一行具有最小的指数，所以矩阵指数位于第四象限。在数学上，矩阵元素通常采用 1-原点索引标志。有一些图像处理软件使用 1-原点索引，尤其是 MATLAB 和 Mathematica，这二者均与数学有很大的渊源。传统的视频和图像处理的扫描顺序通常坚持矩阵惯例，但都是采用零起源索引：行和列从左上角的 $(0, 0)$ 坐标开始算起被编号为 (r, c) 。换句话讲，图像位于 IV 象限（但省略了 y 坐标的负号），但通常使用零起源索引。

数字图像采样结构采用宽×高表示。例如， 1920×1080 系统的列被编号为 0 到 1919，行（史称“图像行”）被编号为 0~1079。

1.3 图像采样

在人类视觉中，三维世界是由眼睛晶状体成像在视网膜上的，视网膜上布满了感光细胞，可以感知大约 $400\sim700\text{nm}$ 波长范围的光。在视频和电影中，我们建立了一个装有透镜和感光设备的相机，用来模拟视觉感知世界的过程。虽然视网膜的形状是部分截取的球面状，但拓扑结构上它是二维的。实际上在相机中，我们没有采用截取的球面，而是采用了平坦的图像平面，如图 1-5 所示。图像科学涉及分析入射在图像平面上的光功率的连续分布。

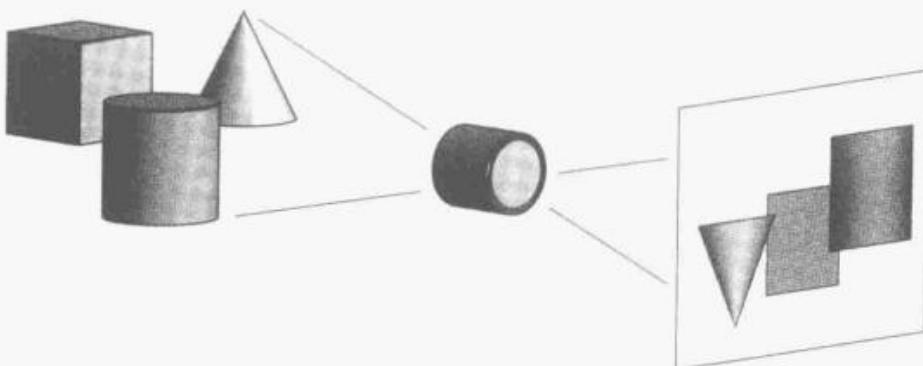


图 1-5 场景、镜头、图像平面

1.4 数字化

物理世界采样的信号通过数字化转换成数字形式，数字化包括两个过程：采样（在时间或空间上）和量化（在幅值上），如图 1-6 所示。采样通常在量化之前，但两个操作可以任意顺序发生。

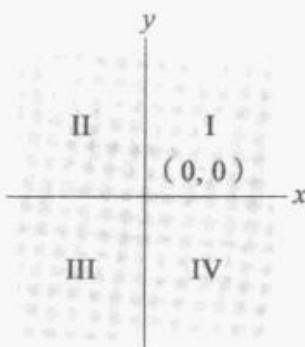


图 1-4 笛卡儿坐标系

笛卡儿坐标系 (x, y) 定义了四个象限，第一象限包含的点拥有正的 x 和 y 值。第一象限的坐标在一些成像系统中被采用，第四象限的点 x 值为正值， y 值为负值。光栅图像坐标点通常表现为随着图像行数增加图像高度下降，也就是说在第四象限，省略了 y 值的负号。

1.4.1 量化

量化是指在某一时刻或空间某点分配一个整数给信号幅度的方法，这将在量化一节解释。实质上所有的图像转换标准（TIFF、JPEG、SD、HD、MPEG、H.264）都涉及像素值，这些像素值并不与场景中或显示的光功率成正比：关于光功率，这些系统中像素值是非线性量化的。

1.4.2 一维采样

一个连续的一维时间函数，比如音频声压级，通过形成一系列的离散值进行采样，每一个离散值是一小段时间上物理量（比如说强度）的分布函数。均匀采样——等时间间隔的均匀采样，几乎是一直采用的方式（详解见滤波与采样一节内容）。

1.4.3 二维采样

一个连续的二维空间函数，通过分配给图像矩阵每个元素一个值进行采样，该值是在一个小的空间区域内的强度分布函数。在数字视频和传统的图像处理中，样本位于规则的矩形网格中。

对于模拟视频，并非水平采样，而是通过扫描且时间上以帧频进行垂直采样。从历史上看，样本也不一定是数字的：电荷耦合器件（CCD）和互补金属氧化物半导体（CMOS）图像传感器都具有固有的采样方法，但它们本身并没有被量化（片上的模/数转换常见于 CMOS 传感器）。但在实践中，一般采样和量化一起进行。

1.5 感知均匀性

如果小扰动编码值在感知量的取值范围内是近似相等可察觉的，那么该感知量可以以感知均匀方式进行编码。考虑收音机的音量控制。如果它物理上是线性的，那么响度感知的对数性质将大部分的感知“动作”控制在取值范围的底部。相反，控制的目的是均匀感知。图 1-7 所示的为带有标准音量锥形调节器的电位器的传递函数。

转动角映射到声压级，这样旋转旋钮 10° 产生一个在控制范围内的近似声音知觉增量。这是许多内建于电子系统工程中的感知考虑的例子之一（另一个例子参见图 1-8）。

相较于线性结构光编码而言，使用模仿人类明度感知的非线性图像编码，能使信噪比性能得到一个动态的提高。理想情况是，在物理光功率下对于分配的编码安排，像素成分值间的步长应与最小可觉差（JND）成比例。CIE 于 1976 年标准化了 L^* 函数，将其作为人的视觉明度敏感性的最佳估计。 L^* 是个有效的幂函数，其指数约为 0.42；在相对亮度中的 18% “中间灰度”对应于 L^* 从 0 到 100 刻度中的约 50。 L^* 的逆函数是一

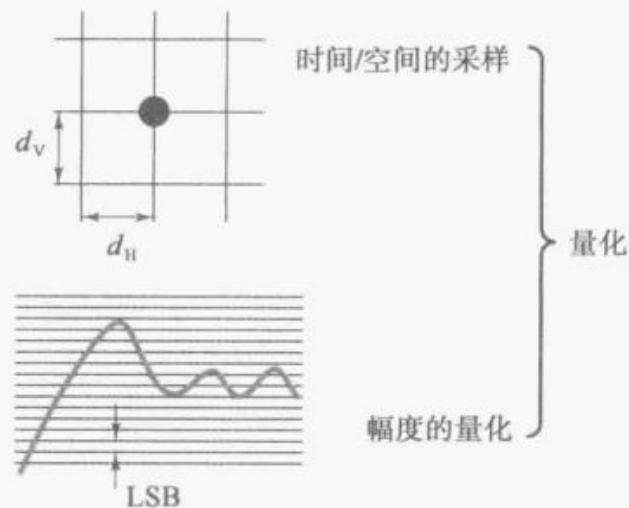


图 1-6 数字化

数字化包括采样和量化，二者可以任意顺序发生。采样密度，单位表示如像素每英寸[⊖]（ppi），与分辨率有关。量化与位每像素数（bpp）或位每通道/分量（bpc）有关。整个数据速率或数据容量取决于这两个因素的乘积

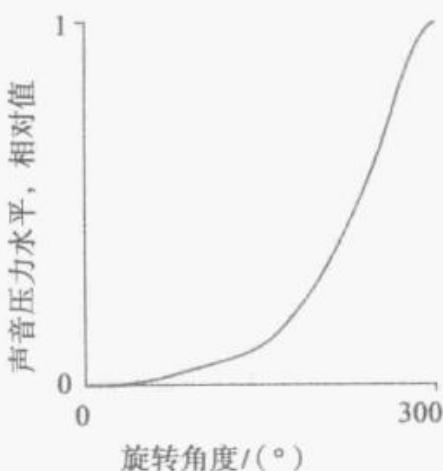


图 1-7 音量锥形调节器

音量锥形调节器在音量调节上采用感知均匀的方式

[⊖] 1 英寸 (1in) = 2.54mm。——编辑注

个指数约为 2.4 的幂函数。大多数商用图像系统集成了一个从数字编码值到线性光亮度的映射，其接近 L^* 的逆函数。

不同的电光转换函数（EOCF）在不同的产业中已经被标准化。

- 在数字影院，DCI/SMPTE 标准化了参考（审批）放映机；这些标准是近似商业电影院的标准。标准的数字影院参考放映机的 EOCF 是一个纯 2.6 次幂函数。
- 在标清和高清中，EOCF 在过去标准化很差甚至几乎没有标准化。只有使用事实上的行业标准 CRT 工作室参考显示器实现了 EOCF 与 2.4 次幂函数的近似一致性。在 2011 年，采用 BT.1886 形式化了 2.4 次幂函数，但是参考白照度和观看条件没有标准化。
- 在高端图形艺术中，使用了行业标准 Adobe RGB1998。该标准建立了参考显示器观看条件。其 EOCF 是一个纯 2.2 次幂函数。
- 在商品桌面计算和低端图形艺术中，使用 sRGB 标准。sRGB 标准建立了一个参考显示器及其观看条件。它的 EOCF 纯粹是一个纯 2.2 次幂函数。

1.6 色彩

为了有益于色彩成像，像素成分表示的是同人的色彩视觉密切相关的数量。在人的视网膜上有三种感光视锥细胞，所以人的视觉有三基色：对于一个正常的观察者，这三色彩不可缺少，且足以表示所有色彩。视杆细胞构成第四感光型，其实现的功能可简单描述为夜视。你看到色彩，是视锥细胞作出了响应。杆状细胞视觉（暗视觉）在几乎所有的彩色成像系统设计中都被忽略了。

通常地，彩色图像用传感器能得到较好的光谱响应度。峰值出现在约 630、540、450nm 处，大致上为红色，绿色和蓝色，其分别对应的光谱带宽约 50、40、30nm。具体细节将在第 25 章、第 26 章中介绍。

在多光谱和高光谱成像中，每个像素有 4 个或多种成分，每一个表示来源不同波段的能量。高光谱系统是指有很多频谱分量的设备。目前，对多少成分构成多光谱或高光谱，还没有一个被广泛接受的定义。我定义，多光谱系统为具有 4 到 10 个频谱分量的系统，高光谱系统具有 11 个或更多分量的系统。高光谱系统可以描述为虽然具有色彩，却是为科学的目的而设计出来的，而不是为了视觉赏析：在高光谱系统中一组像素分量值通常与色彩感知没有太大的关系。除了一些高度专业化的应用，例如，卫星成像和艺术品的保存或仿制，多光谱和高光谱的技术在商业成像中并不常见。

1.7 亮度和色差分量

有些数字视频设备直接使用 $R'G'B'$ 分量。但是人视力在颜色信息方面的感知能力比明度感知能力要弱许多。倘若保留着非彩色的细节，彩色的细节就可以通过亚采样减少，这是一种空间滤波（或叫做平均滤波）。

6R17

Snowfield
Champ de neige
00NN 72/000

72

Vell
Voids
00NN 53/000

53

Granite Grey
Gris Granit
00NN 37/000

37

Seal Grey
Gris phoque
00NN 25/000

25

Grey Tabby
Chat mouche
00NN 16/000

16

Deep Onyx
Onyx foncé
00NN 07/000

7



图 1-8 灰色绘图样点表现出感知的均匀性

灰色绘图样点表现出感知的均匀性。制造商的目的是使反射率涵盖相当广泛的范围，以致人做视觉判断时，样点是均匀间隔的。每个芯片的制造商编码通常包括一个近似的 L^* 值。在图像编码中，我们使用一个相似的方案，用编码值（像素） V 代替 L^* ，并且用数量为 100 或 1000 个编码而不是 6 个

从事色彩研究的科学家可能会通过采取用相对亮度作为一组 RGB 三色值的加权和的方法进行亚采样，然后通过一个非线性的转换方程近似求出 CIE 明度的值 (L^*)。在视频图像领域，我们摒弃了色彩科学家的理论，而采用工程近似法，这部分内容将会在恒定亮度这一章节进行详细说明。简单地说，分量视频系统所输送图像数据包括了亮度分量 Y' ，用来近似等于明度和编码非彩色分量，以及两个色彩的不同分量——在传统模拟领域的 P_B 和 P_R ，数字系统中的 C_B 和 C_R ——它们只反映了色彩度而忽视了明度。色差分量通过进行亚采样（水平采样，或者水平、竖直采样同时进行）来降低数据传输速率。 $Y'C_B C_R$ 和 $Y'P_B P_R$ 将在第 12 章中详加解释。

1.8 数字图像的表示方法

静态数码图像和动态序列可用许多不同的文件、内存和流媒体格式用来传输。大多数情况下每个像素的格式都有 3 个分量值（来描绘附加的红、绿和蓝三种颜色分量）。在消费级别的电子和商品计算中，大多数格式每个分量会有 8 位。在专业应用，如在视频工作室和数字影院的应用中，每个分量上有 10 位、12 位，乃至更多位数也是正常的。

成像系统通常用于优化人类感知的其他方面，举例来说：JPEG 和 MPEG 压缩系统利用视觉的空间频率特性进行压缩。这样的优化通常称为感知编码。

实际中，几乎全部商业成像系统都使用感知编码，即像素值按照接近人视觉的灰度级辨别能力水平排列。在色彩学中，大写字母 R 、 G 、 B 用来表示三色值，其值的大小和三个波段上光的能量成比例。三色值的感知是不均匀的。不管是显式的还是隐式的，在所有商业数字图像处理系统中，当需要显示的 RGB 三色值的幂次上升到大约 $1/2.2$ （也就是大约 0.45）和 $1/2.6$ （也就是大约 0.38）之间时，像素分量值就会进行编码。这样构造出来的像素值用加上撇号的字母表示： $R'G'B'$ （尽管上撇号经常省略，以免引起混乱）。

为了对图像数据进行合理可靠地交换和解释，数字图像标准定义了参考黑色和参考白色像素值。数字图像标准通常为参考白色指定一个目标照度，大多数数字图像标准没有为参考黑色提供特定的反射系数或者相对照度，这暗示我们的显示系统将参考黑色设置成尽可能的黑。

在计算过程中，8 位的数字图像数据的范围是从编码为 0 的参考黑色到编码为 255 的参考白色。sRGB 标准要求在显示时指数 (γ , gamma) 要达到 2.2。参考白色显示时亮度则应该达到 $80\text{cd} \cdot \text{m}^{-2}$ ，但实际中的值达到 $320\text{cd} \cdot \text{m}^{-2}$ 都是正常的。

在消费级的数码视频中，图像数据都将参考黑色设置为 16，将参考白色设置为 235 的 8 位分量进行编码（在图像工作室中，一般是 10 位编码）。图像工作室要求显示时的指数要达到 2.4。通常的做法是在黑暗的观测环境（约 1lx）和黑暗的周围环境（约 1%）下将参考白色的亮度设置成 $100\text{cd} \cdot \text{m}^{-2}$ ，当满足这些条件时，就能在消费者所处的环境中可靠地显示。消费类显示器比 $100\text{cd} \cdot \text{m}^{-2}$ 更明亮是正常的；消费者的观看环境经常是相当明亮的（大约 100lx），周围的环境往往只不过是昏暗的（5%）。由于消费类显示器和观看条件不同于图像工作室，保持画面的外观就需要进行显示调整。

交互式数字视频显示 (DVI) 系统，定义为能承载 $R'G'B'$ 数值范围为 0~255 的计算机显示系统。高清晰度多媒体接口 (HDMI) 和显示端口接口也同时兼容以上编码，不过这里的 $R'G'B'$ 值范围是 16~235，能够容纳比 8 位更宽的位宽，而且也同时兼容 $YC_B C_R$ 编码。串行数字接口 (SDI) 和高清串行数字接口 (HDSDI) 都能够容纳 10 位位宽的 $R'G'B'$ 或者 $YC_B C_R$ 编码，它们的范围为 64~940（即 16×4 , 219×4 ）。

平版印刷的图像数据通常定义为 8 位的编码，编码值代表每个像素位置的青色，品红，黄色和黑色（CMYK）的数量，如（0, 0, 0, 0）代表没有显示（显示基板的颜色，通常是白色），（0, 0, 0, 255）则代表黑色。感知均匀性问题通常使用打印机中网点扩大的概念来表示。用户接口通常以 0~100 来表示 0~255 的范围，换而言之，这时的接口值表示的是百分比。

1.9 标清和高清

1995 年以前，“电视”一词专指视频格式为 480i（历史上表示为 525/59.94，或称“NTSC”）或者 576i（表示成 625/50，或称“PAL”）的电视标准。宽屏电视，高清晰度电视（HDTV）等其他新电视系统的兴起，为电视这个术语引入了更多的不确定性。我们以前所说的电视现在应该称为标清电视（SDTV）。广泛使用的视频技术正逐步将娱乐节目和传统意义上的电视分开，而应用软件又如此丰富多彩，所以我往往将“电视”一词忽略，而直接使用标清电视（SD）或者高清电视（HD）。

作者认为 480p、540p、576p 都不应算作高清，但 720p 属于高清（这里对清晰度的定义标准是几何意义上的 540 和 720）。

令人惊奇的是，目前仍然没有广泛接受的标清和高清的定义。按照我的分类标准，任何满足帧频 23.976Hz 的视频系统，或者是列数少于 1152，行数少于 648，也就是总像素少于 3/4 百万像素的数字图像系统，都可以称为标清系统（SD）。同时，屏幕比例为 16:9，帧频达到或多于 23.976Hz，列数多于 1152，行数多于 648，也就是像素点总数多于 72.9 万像素（约（3/4）百万像素）都可以称为高清系统。

面积为 $1m^2$ 的 16:9 的图像，其具体尺寸是 $(4/3) m$ 乘以 $(3/4) m$ （约为 $1.33m$ 乘以 $0.75m$ ），对角线长度约为 $1.53m$ (60in)。这一尺寸代表了实际消费者最大直接观看显示范围。宽度为 $1m$ 的 16:9 屏幕，高度约为 $0.5625m$ ，视频区域面积约为 $0.5625m^2$ ，其对角线长度约为 $1.15m$ (即 45in)。这一尺寸大致代表了目前消费者购买的高清电视的中档水平。对于 1920 像素乘以 1080 像素分辨率的屏幕来说，其像素尺寸约为 $1/1920m$ ，或者说 $520\mu m$ 。每一个 RGB “子像素” 尺寸约为 $174\mu m$ 乘以 $520\mu m$ 。

1.10 方形采样

在现代数字成像，包括计算、数字摄影、图形艺术、高清视频系统中，采样点（即像素）通常在水平方向和垂直方向均匀分布——它们在水平和垂直方向采样密度相同。这些系统采用方形（“方形像素点”）采样，它们有相同的采样幅型比（SAR）。通过使用方形采样，图像的列数与图像行数的比值即为采样幅型比。

这里的方形一词指的是采样的方法，并不是指图像信息均匀分布在与每个像素相关的正方形区域，一些 1080i 的高清压缩系统像素幅型比达到 $4/3$ 或 $3/2$ 。

传统的成像和包括数字标清系统在内的视频系统水平和垂直采样点不等（非方形采样）。这种情况有时会被误认为是“矩形采样”，但是方形也是一种矩形。在 20 世纪 90 年代早期，激烈的讨论推动了高清系统采用方形采样。1995 年纽约时报写道：

高清电视信号可以采用多种不同的数据格式，不仅仅依赖于逐行信号或隔行扫描信号，而且依赖于诸如与圆形像素相对应的方形像素……

技术人员可能觉得上述声明很滑稽。然而令人吃惊的是，或者说，在无意之间，它也揭示了一些技术上的真谛：在对连续图像进行采样时，每个像素点关联的图像信息扩散到

相邻像素点上，并且，在理想条件下它们是圆对称的。

1.11 幅型比的比较

在 20 世纪 90 年代中期，高清视频正在经历标准化，标清和高清在对图 1-9 所示的多方面进行了对比，其依据是 4:3 和 16:9 不同的幅型比。这些对比在相同的高度下进行。

图 1-9 所示的是分别基于相同的高度、宽度、对角线、面积，对传统电视（SD）和高清图像幅型比进行比较。相对于标清电视，高清电视根本性优势是有更多的像素数量，而以上的比较是忽略这一优势的。正确的比较是基于相同的图像细节，一个像素点的角弦应当被保存起来。

高清晰度（或分辨率）并不能将 6 倍于像素点的内容挤压到相同的视角里，因此以上所有方面忽略了高清电视自身的优势。取而代之的是，保持一个像素点的角弦，进而整个图像就能占据观察者视野中更大的区域。高清可以大大增加画面视角，正确的比较标清电视和高清电视并不是基于图像幅型比，而是根据画面细节。

1.12 幅型比

随着高清家用电视接收机的出现，在 16:9 (HD) 的显示器上显示 4:3 (SD) 的图像和在 4:3 (SD) 的显示器上显示 16:9 (HD) 的图像成为急待解决的问题。在高清电视标准化过程中，不完全是开玩笑的讲，我曾提议应将 4:3 格式的标清图像按像素映射到高清帧内，并保持图像幅型比和相同的图像细节，如图 1-10 所示。我希望相关规定被指定，用来保证消费者能够放大标清图像——但消费者将会意识到两种截然不同的图像来源（我的想法没有被采纳）。

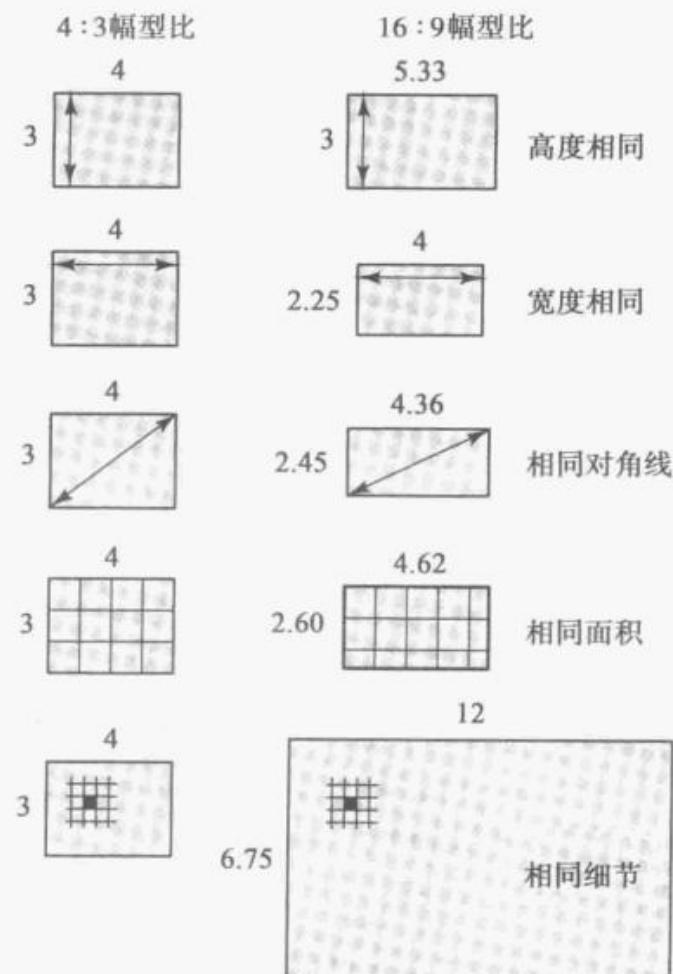


图 1-9 图像幅型比进行比较

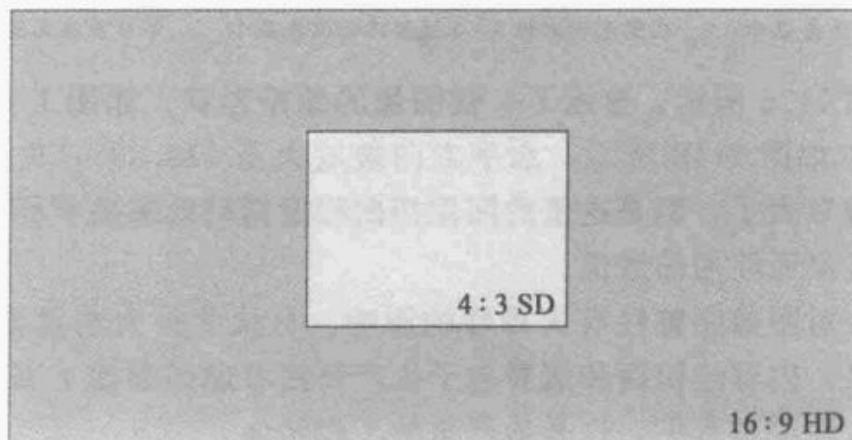


图 1-10 标清到高清像素映射

裁剪图像的宽度可更改图像大小，可使宽屏 16:9 图像调整为 4:3 格式图像；但是，图片内容会丢失，同时原有的创新点也会被损坏。图 1-11 和图 1-12 显示了用中心裁剪方法调整 16:9 格式图像的结果，图像的细节可能会受损。



图 1-11 幅型比的变化

可能会损害创作意图，考虑这个 1.78:1 幅型比的帧，这两个人物在冒险之前，先观察水面



图 1-12 按中心切割方式处理 4:3 幅型比的图片

当按中心切割方式处理 4:3 幅型比的图片时，一个人被剪切掉了；整幅图片内容被改变。许多戏剧性和喜剧效果都取决于图片边缘的动作来表现

在下一页面的顶部，如图 1-13 所示，平移和扫描的作用是按场景选择 4:3 格式图像中要保留的区域，进而缓和因裁剪所引起对原图像的内容损失。

许多导演和制片人拒绝他们的影片因裁减而改变；因此，许多 DVD 中影片以 Letterbox 方式发布，如图 1-14 所示。在 Letterbox 方式中，宽屏图像被完整保存下来，4:3 框架的顶部和底部没有被用到（通常，没用到的部分显示为灰色或黑色）。

传统的 4:3 节目源可以以 Pillarbox 方式在 16:9 设备显示，如图 1-15 所示，显示屏的高被全部利用；宽屏框架的左右两边空余。但是，消费电子（consumer electronics, CE）厂家担忧消费者抱怨标清上变频后屏幕中存在未使用区域。所以，消费电子厂商制定了扩张水平方向图像方案，以消除左右两侧的空余。



图 1-13 平移和扫描

裁剪宽屏片源的宽度，将 16:9 的节目源在 4:3 幅型比的设备显示

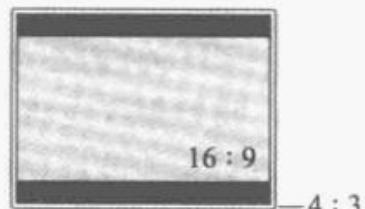


图 1-14 Letterbox 方式

显示宽屏节目源，将 16:9 的节目源在宽度相同的 4:3 幅型比的设备显示

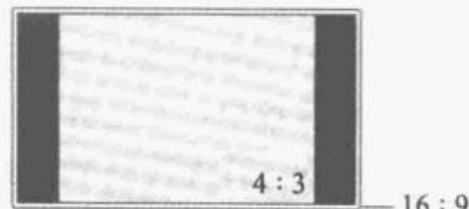


图 1-15 Pillarbox 方式

有时叫做 sidebar，将窄幅型比的节目源在高度相同的 16:9 的设备显示

中心图像，如图 1-16 所示，显示了一幅图像的原始形状。如图 1-17 所示，图像水平方向被压缩至 75%。如图 1-18 所示，水平方向被放大至 133.3%，失真非常明显，以致你也许会怀疑结果被夸大了，但是这里的图像扭曲程度精确地取决于标清-高清和高清-标清转换中以适应帧宽度而使用的数值。

这些缩小与扩展对图像完整性有灾难性的影响。自从北美为消费者引入高清电视后，这些情况已司空见惯。内容经销商和消费电子生产商在考虑幅型比上面的失败，就个人观点，在高清系统引入北美过程中，这是最严重的工程错误。



图 1-16 一张 Barbara Morris 的正常图像



图 1-17 挤压图像

如果 16:9 的节目源要强行改变尺寸适应 4:3 帧，挤压至 3/4 很必要



图 1-18 拉伸图像

如果 4:3 的节目源要强行改变尺寸适应 16:9，拉伸至 4/3 很必要

1.13 帧频

标清广播电视一直以来使用隔行扫描方式。在 480i (NTSC) 系统中，规定帧频为 30/1.001Hz (29.97Hz)；在 576i (PAL) 系统中，规定帧频为 25Hz。复合 PAL 制与 NTSC 制视频帧频是固定的。分量视频系统在帧频的选择上具有灵活性。但是，生产者和分频设施通常会将其锁定在一、两个帧频上，25Hz 或 29.97Hz 对于频段的国际分配来说，帧频转换在分频设施与消费者设备处均很重要。

帧频曾经以区域为基础划分，以配合现行交流 (AC) 电源线频率。20世纪 90 年代，曾致力于为高清建立一个世界统一的帧频；但努力并未成功。高清的产生与传播源于现行的电源线频率，50Hz 或（名义上的）60Hz。某些照明装置以 2 倍于交流 (AC) 电源线频率（尽管远高于感知闪烁敏感性）采样闪现。如果一台摄像机工作于不同于电源线频率的帧频上，这样的闪烁就会与摄像机的帧频产生共振，产生一个阻止的低频频闪。

从商业 IP 网络到消费者个人计算机的视频分发，通过解码来恢复节目的本地帧频，但通常没有方法对显示系统进行同步，所以经常导致不好的动态效果。

第 2 章

图像的采样和显示

数字图像的基本概念总结如图 2-1 所示。图像数据是经过捕获，处理或者是记录，然后呈现给观众的，图 2-1 进行了大概的描述，稍后将进行详细的展开，图像的显示结果是由显示和观察条件决定的。观看图片的条件通常是不同于场景捕获的时候，如果这些条件不同，必须施加图像渲染，获取图片的可信的映射，以达到对场外观众真实的显示（而不是它的物理激励）。

图 2-1 显示了图像获取过程，通过照相机获取图片，照相机通过获取场景的光亮，将光亮转换为信号，在大多数照相机中执行特定的图像处理操作。然后信号被记录，进一步处理或者发行。最后，该信号在显示装备中被转换为光。所显示图像的表现状况依靠显示条件（比如说峰值照度）；观看条件（比如说显示器周围的环境）和显示器，以及它的配置（比如说对比度）的共同影响。有时候场景比所显示的图像更加明亮：

在白天获取场景的亮度白色值为 $30\,000\text{cd} \cdot \text{m}^{-2}$ ，但是摄影棚产生的亮度白色值仅仅为 $100\text{cd} \cdot \text{m}^{-2}$ 。通常成像目标与场景的物理激励不匹配（比如说，白天的照度等级），而与观众对场景的期望相契合。产生一个场景匹配要求施加一个不一般的映射——称为图像渲染——映射场景照度到显示照度。

在阳光明媚的中午观察花园里的花朵，日落半个小时后观察同样的花园。物理上讲，除了有较低的照度水平，花的光谱没有任何变化。然而，花朵在日落之后明显不再那么的丰富多彩：照度减弱，色彩也会衰退，这是以著名色彩科学家亨特（R. W. G. Hunt）命名的亨特效果。图像通常是显示原场景的一小部分，或许图像照度只是采样时照度的百分之一或者千分之一。如果呈现图像的照度正比于场景的照度，展现的图片与原来的场景相比色彩度会减小，对比度会降低。

为了使呈现的对比度和色彩与原来的场景相媲美，我们必须改变图像的特性。科学家和物理学家在成像系统中努力实现数学上的线性变换，然而要求的变化引起显示相对照度不同于原来场景照度的比例。困境是这样的：我们可以实现数学上的线性度，或者实现正确的外观，但是我们不能同时实现二者！成功的商品成像系统是以牺牲数学运算来达到正确的感知结果的。

2.1 图像状态

在许多专业的成像应用中，图像在上映之前会进行审查和批准。尽管图像数据来源于原来场景的比色度，任何技术或者是创造性的决定都会导致图像数据与原来链接的变更。

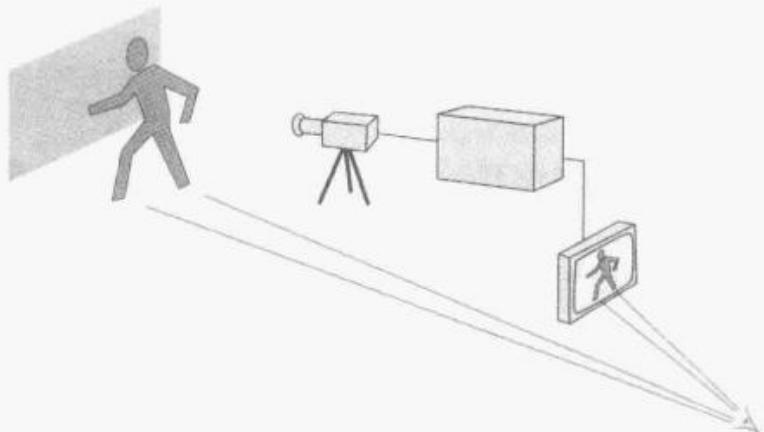


图 2-1 图像获取

想一想电影欢乐谷，颜色被用做讲故事的设备。故事立足于在灰度和色彩描绘上对人物的刻画（见图 2-2），最终电影的 $R'G'B'$ 值不能准确代表摄像机前的画面！尽管这个例子来源于娱乐业，但是出于审美目的而调整颜色的例子很多。

图像渲染通常是非线性操作，用一个方程或者是一组方程不能很容易描述。一旦图像渲染被执行，它的参数通常不会被保持。在许多图像应用领域，图像数据被控制，以达到艺术效果——比如，婚礼照片的颜色可能被摄影师有选择的改变。在这种情况下，考虑图像渲染的数据可能和整个原始影像一样复杂。

图像系统的设计决定图像渲染适用的地方：

- 在消费类数码摄影和视频制作中，画面渲染通常应用在照相机里。
- 在电影制作中，图像补偿通常应用在处理链上。

如果一个成像系统具有一个从场景照度到图像编码值直接确定性的链接，在色彩管理术语中图像数据称为有一个场景关联的图像状态。如果从图像编码值到显示设备显示照度有一个直接确定性的链接，这个图像数据就称为显示关联。

视频标准，例如，BT. 709 和 SMPTE ST274（稍后将进行详细解释）的图像状况最好情况是不清楚，最坏的结果是报错。因此，视频工程师总是错误地认为，视频数据是按色度链接到场景的。数码相机用户可能会认为他们的相机捕捉到了“科学”，当捕捉到 TIFF 或者 JPEG 图像时，用相机算法进行渲染，因此与现场场景的色度链路被破坏了。在这些应用中最重要的不是光电转换函数（OECF），它曾经用来使光亮度从现实场景中映射到图像数据值中，而是 EOCF 被期望于将图像数据值映射为呈现给观众的光亮度。

2.2 EOCF 标准

在成像系统中，原始的图像数据受到检查和审核，真实地呈现原始图片需要从图像数据到光亮的连续映射——在娱乐性应用中，从音频信号到声音——从原始的环境到最终的视觉环境之间。

图 2-3 描述了起源、批准、发布和呈现的基本链。起源被描述为一个“黑盒子”，“从图像数据到显示光亮的映射”涉及一个电光转换函数（Electro-Optical Conversion Function, EOCF）。从草图中可以清楚地知道真实的再现原景需要在审核显示和最终呈现时对 EOCF 进行匹配。因此 EOCF 存在于任何图像转换标准中（明显的或隐含的）。真实的再现同样要求参考观看条件的协议（明显的或隐含的）。

为了充分的利用有限的信道承载能力，在商业应用中的电光转换函数结合了均匀感知，我们即将讨论这个话题。



图 2-2 作为一种戏剧化设备的色彩

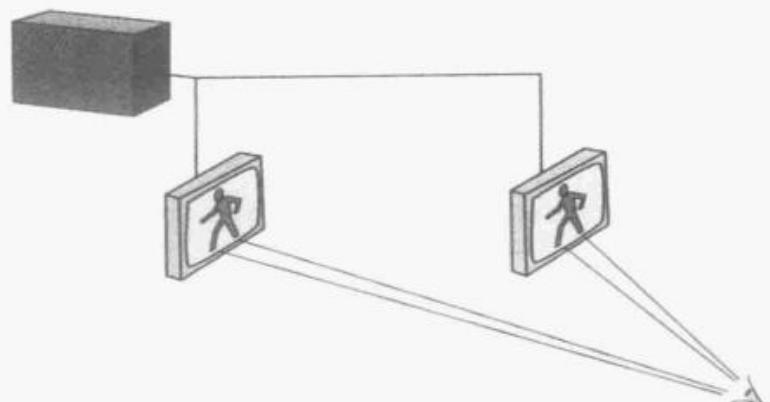


图 2-3 图像审核

2.3 娱乐节目

娱乐是代表了一种具有经济意义的重要应用，它特别值得一提。数字视频、高清电视和数字电影都涉及采样、记录、处理、发布和演示的过程。本文将采用一般性的语言“节目”作为一个电影、电视节目，或者一个商业广告片段的简称。图 2-4 表达了整个链的草图。



图 2-4 生产流程

产品的整个过程涉及获取、记录和处理。在实景电影中，产品的制作仅限于图像的获取（布景或者外景），接下来的就是后期制作（“post”）。对于视觉元素都是数字的电影来说，后期制作就涉及数字中间处理或者是数字图像（DI）。

产品最终阶段就是摄影棚展示作品的审核和显示，或者在数字影院，发行的电影则在影院的投影机下放映（如果发行涉及压缩，那么审核的过程适当地涉及在演播室的压缩以及使用解压工具解压缩）。审核之后，就是处理，打包和发行。

专业的视频制造者很少会站在观众的角度去寻求电影场景的准确表达。除了纪录片制作者，电影制造者经常进行创造性的选择来改变现实。他们希望当电影通过发行链完成它的整个过程，最终呈现给观众的不是原来的场景，而是在摄影棚中展现在导演面前的经由他本人批准的后期制作的最终产品。在色彩管理方面，电影和视频图像数据是以显示为参考依据的。

2.4 图像获取

人们使用相机获取一个场景的图像数据，期望图片放映时，它能近似真实地再现这个场景。室外白色的亮度能够达到 $30\,000\text{cd} \cdot \text{m}^{-2}$ ，但是很难找到一个电子显示设备的照度能够达到 $450\text{cd} \cdot \text{m}^{-2}$ ，专业的高清电视机能够满足的参考白色值标准是 $100\text{cd} \cdot \text{m}^{-2}$ 左右。场景照度线性变换到显示屏——实际上采用 0.015 或者 0.01 的因子缩放绝对照度——不能展现与外面场景相同的效果。人们使用相机希望最终在显示屏上能够有外观上的近似匹配。因此图像渲染必须要进行。在高清电视机和消费类静态摄影中，渲染是在相机中进行的，在数字影院和专业的（“原生态”）静态摄影中，渲染是在后期制作中进行的。

有些人用“现场到屏幕”来描述这样的目标，即把描述场景的照度和颜色的呈现精确地传输到显示屏。

除非适当考虑显示现象——就是说，除非强加图片渲染——要不然这种效果注定是要失败的。

2.5 面向消费者的视频获取

图 2-5 总结了面向消费者的视频获取。消费者可以通过信号处理对获取资源实行创意方面的控制；然而，图像渲染是由相机中的算法施加的。摄像机设计成可以进行信号

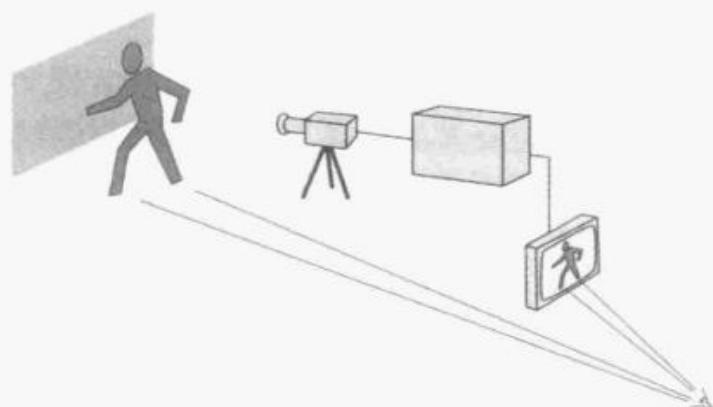


图 2-5 面向消费者的视频获取

编码的，这些信号能在消费者的客厅中展示。演播室是建立在假设观看亮度为 $100\text{cd} \cdot \text{m}^{-2}$ ，而周围是黑暗的环境（现在，大约是 1%）。消费级的摄像机基于可比参数进行图像渲染。

2.6 消费类电子显示器

在消费电子领域，有多样化的显示设备（有不同的幅型比，不同的照度峰值，不同的色域）并且有多种观看环境（有的明亮，有的黑暗；周围环境有的明亮，有的灰暗，有的黑暗）。

不同的消费类显示设备有不同的默认 EOCF。每一个产品的 EOCF 都是在工厂中实现的，通过采用一个合适的方法去适应那个产品所期望的观看条件。传统领域的电视接收机的 EOCF 接近于演播室中的 2.4 次幂函数；然而现代的消费类接收机明亮度要比 $100\text{cd} \cdot \text{m}^{-2}$ 亮得多（今天最高达到 $350\text{cd} \cdot \text{m}^{-2}$ ），然而这个更高的明亮度要求伽马值要低一点，现代典型的值为 2.1~2.3。一个家庭影院投影机在一个相当黑的环境中的效果可以与演播室的参考显示器相媲美（参见第 37 章），通常都会有大约 2.4 的“伽马”值效果。一个 PC 的默认的 sRGB EOCF 值是 2.2 的伽马值。

消费类电视接收机厂商通常强行进行信号处理，自称为改进图像——经常用“清爽自然”或者“生动形象”等形容词来形容。然而，导演可能会有充分的理由认为图片看起来不自然，暗淡的，或者嘈杂！创造性控制可以出现在制作中，而不是在展示中。主创人员通常轻视消费类的处理过程，例如数字现实创造（DRC）或者数字自然影像技术（DNIe）。

第 3 章

线性光和感知均匀性

每一个灰度图像的像素值代表了广义上的明亮度。然而明亮度是根据该区域散发光的多少被正式定义为视觉感官的属性。这种定义显然是很主观的：明亮度不能被测量，因此不能作为数据图像的度量。而且，根据色彩表现理论，明亮度没有最大值：明亮度跟任何值没有关系。

光强 (intensity) 是某个方向上的辐射功率，也就是每单位立体角的功率 [$\text{W} \cdot \text{sr}^{-2}$]；辐射 (radiance) 是指每单位投影面积的光强度。这些术语都忽略了波长因素，但是在彩色图像中，波长是非常重要的！这些都不适合作为彩色图像数据的度量单位。

参阅附录 B 有关于辐射度和光度的介绍。声音强度与光强在概念上有很大的不同。

照度是指光中与视觉明度感知相关的光谱灵敏度的加权值，照度与光强度成正比；在 SI 系统中，它代表了每平方米的坎德拉量 [$\text{cd} \cdot \text{m}^{-2}$]，通常称为 nit [nt]，图像系统中很少采用正比于照度的像素值，而经常采用与照度非线性相关的数值。

照度描述了落在物体上的光，从技术上讲，是集成在一个半球上的照度。

明度是被 CIE 定义为某个区域的明亮程度，与被照亮区域的明度相似，其表现为白色或者强放射性。纯化论者可能会说这个判定方法太主观了。然而，客观量 L^* 定义为相对照度感观反应的标准估计。它是用相对照度的非线性特征方程模拟在特定情况下的视觉反应来计算的。一些灰度级成像系统中有正比于 L^* 的像素值。

测量明度的值大致相当于 CIE L^* （通常在 0~10，不是 100）。在图像科学领域，与标准图像一致的情况很少。

遗憾的是，有很多从事于数字图像处理和计算机图形处理的人都对这些条款视而不见。在 HSB, HSI, HSL 和 HSV 系统中 B 代表着亮度 (brightness)， I 代表着强度 (intensity)， L 代表着明度 (lightness)， V 代表着数值；然而，这里没有一个与色彩科学领域公认的明亮度、强度、照度或者数值相关。

彩色图像是以三基色值来感应和传播的，它的幅度正比于强度，其光谱成分是根据彩色图像领域精心挑选的。相对照度可以看做是一个非常有用的三基色值，除了那些特点，三基色值是以三个为一组的。

数码相机传感器产生的值，正比于辐射强度，接近于红绿蓝 (RGB) 三色值。我们把这些值称为线性光。然而，在大部分成像系统，RGB 三色值采用非线性编码——伽马修正——来模仿人类的视觉系统。大部分图像编码系统使用的 $R'G'B'$ 值并不正比于强度，在这里的标记符号表示施加的感知驱动的非线性参数。

亮度信号（“Y'”）是 $R'G'B'$ 的加权和，它是视频、MPEG、JPEG 和类似图像编码系统的亮度/色彩差分编码的基础。在视频中，伽马修正中的非线性变换形成 $R'G'B'$ 分量随后融入到亮度和色度 ($Y'C_B C_R$) 分量中。

3.1 对比

在图像应用领域，对比度一词承载过多含义。“对比度”涉及一个场景的物理性能，

光在图像中的物理特性，两者之间图像系统映射特性，显示系统（任何独立图像）的物理特性，或者关于感知的各种属性。

摄影师经常说场景具有“高的对比度”或者“低的对比度”。一个场景的对比度可以用其照度的变化系数来形容——就是由平均值来计算的照度的标准偏差。场景的对比度与关于“重心”的照度直方图密切相关。一个黑暗的场景与一个明亮的场景可能会有一个相同的对比度。场景对比度不受噪声的影响。

摄影师也经常说高色调和低色调场景。粗略地讲，这些主要是分别讲亮或暗。从记录照度的对数累计直方图上，色调可以被量化为沿着 x 轴的在第 10 到第 90 个百分位的分位数，从第五十个百分点处开始下降。

3.2 对比度

一个显示器的基本属性是它的对比度，也就是指定的亮度与暗度的比例——通常是指照度参考白色和参考黑色。图像间（或者开关，或者连续）的对比度通过单独列出的全白和全黑图像来测量。图像内的对比度是通过测量如由 ANSI IT7.228-1997 规定的（现在已经撤回，由于投影仪）或者 ITU-R BT.815 规定的，单个测试图像中的白色区域和黑色区域而得到的。显示系统的视觉性能的最好特征是图像内的对比度。

BT.815 规定了拥有 16:9 幅型比的对比度测试图像，这对测试对比度非常有用。如图 3-1 所示，对比度是由黑色方形的平均照度除以白色方形的照度得来。

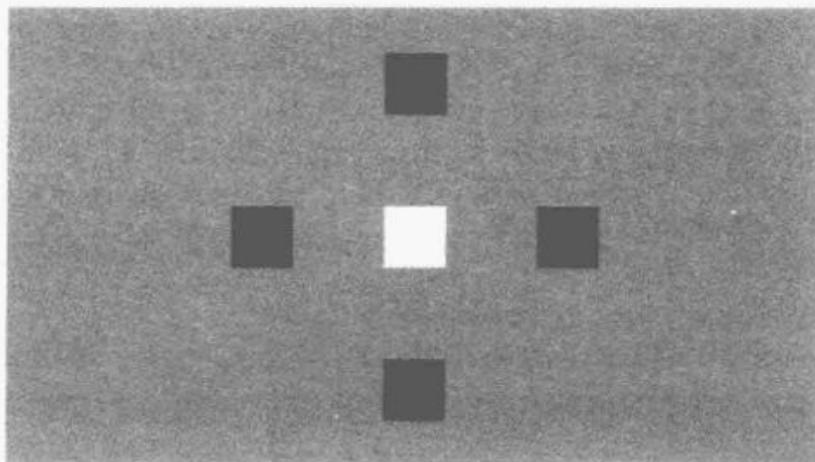


图 3-1 ITU-R BT.815 模式

用来表征图像内对比度，灰色的背景是 8 位的接口代码 126。参考白色的小方块位于图像的中间；它占据了图像高度的 $2/15$ 。四个相同的参考黑色小方块环绕着中心，在图像高度的 $1/3$ 偏置处，分别位于 12 点钟、3 点钟、6 点钟、9 点钟方向。

在实际成像系统里，许多因素会联合起来增加黑色的照度，从而减小了对比度，破坏了图像的质量。在一个电子显示或者投影图像中，图像内对比度，通常小于 800:1，这是因为显示系统闪光和周围环境溢出的杂光引起的。典型的对比度如表 3-1 所示。对比度是一个主观图像质量的决定性因素，以至于一个被更高对比度再现的图像可以比一张具有更高测量空间频率内容的图像看得更清楚。

因为对比度只涉及两个量的测量——参考黑色和参考白色——它是独立的传递函数。对比度（和它的第一个表亲，动态范围）对位深度没有任何介绍。一种二级的图像包含了参考白色和参考黑色，并且足够产生测量对比度所需要的光刺激。

表 3-1 典型对比度总结

应 用	最大(参考)照度 $(cd \cdot m^{-2})$	典型 PDR 照度 $(cd \cdot m^{-2})$	典型图像间 对比度	典型 BT. 815 图像 间对比度	最小 L^*
影院(投影)	48 (48) 120	24	1000 : 1	100 : 1	9
高清, 专业演播室	(100)	90	10 000 : 1	1000 : 1	0.9
高清, 客厅(典型)	200	200	1000 : 1	400 : 1	2.3
办公室(sRGB, 典型)	320	320	100 : 1	100 : 1	16

3.3 感知均匀性

如果一个编码值的小扰动在这个值范围内是近似等价感知的，则感知量的编码是均匀的。如果两个照度的比例小于 1.01，人眼是不能区分亮度水平的——换句话说，照度差的视觉阈值大约是 1%。这种对比灵敏度阈值是由一种测量模式建立的，比如图 3-2 所示的草图。详细的介绍将会 23.3 节中展开。理想情况下，像素值会被放置在沿着从参考黑色到参考白色的恰好辨识差(JND)递增的地方。

图 3-2 所示的是对比敏感度测试模式显示，在 ΔL 超过 L 的百分之一时，不同的照度在一定的条件下可以观察到，这个阈值称为恰好辨识差 (JND)。

3.4 “编码 100” 难题和非线性图像编码

考虑 8 位像素值正比于照度，零值代表黑色，最大编码值 255 代表白色，如图 3-3 所示。编码 100 位于某个尺度的点上，这个尺度是，相邻的照度值比率是 1%。由于视力的对比度阈值是 1%，编码 100 和 101 的采样的区域的边界是可以看得到的。

像素值降低到小于 100 时，相邻代码的不同亮度的差别越来越容易感觉到：在编码 20 中，相邻的照度值的比率达到 5%。在灰色阴影平滑变化的大区域中，这些照度的差异很可能是可见的甚至是令人讨厌的，可见的照度跳跃会产生轮廓或者带状伪影。

高于 100 的线性光编码，没有带状伪影，然而随着码值的增加逐渐接近白色，编码减小了感官的效应；编码 200 时，相邻码值之间的照度比值仅为 0.5%，低于视觉阈值。编码 200 到 201 在视觉上没有什么区别，编码 201 在不被察觉的情况下可以被丢弃。

高品质的图像显示要求白色到黑色照度的对比度至少为 30 比 1。在 8 位线性光编码中，最亮照度(编码 255)与最暗照度的比率在没有光带(编码 100)的情况下仅是 2.55 比 1。因此 8 位的线性光编码不适用于要求 30 比 1 比率的高质量图像。

缓解“编码 100”问题的方法是将尺度上端点放在码值远远高于 100 的地方，如图 3-4 所示。如果亮度由 12 位的数据所表示，白色编码为 4095，编码 100 和白色之间的照度比达到 40.95 : 1。然而 4096 中大量的亮度值不能被视觉区分；比如码值 4000 到 4039 在视觉上不能区分。这样我们可以用少得多的代码值，并且根据光功率非线性将他们放在一个感官的尺度上，而不是使用大量的位值对照度线性编码。这种非线性光——或者感知均匀性——在数码影像中的编码是很普遍的；它运作的这样完美，许多从业人士甚至不知道它已经发生了。

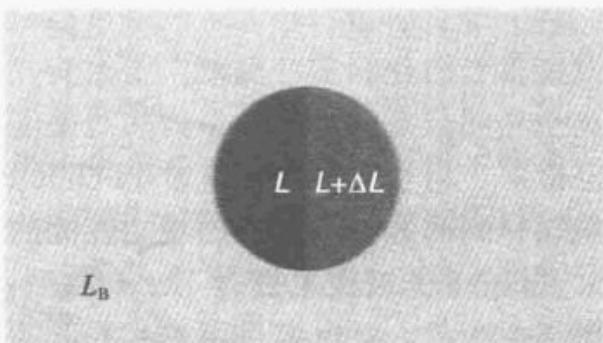


图 3-2 对比灵敏度测试

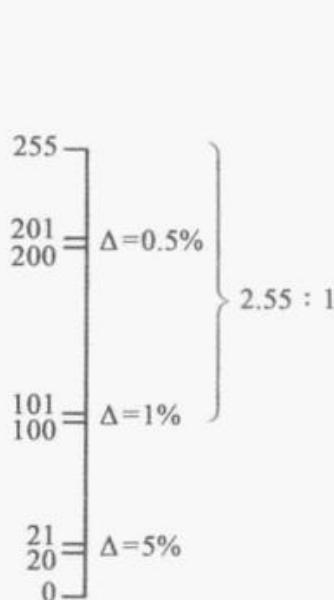


图 3-3 “编码 100” 线性光编码的问题

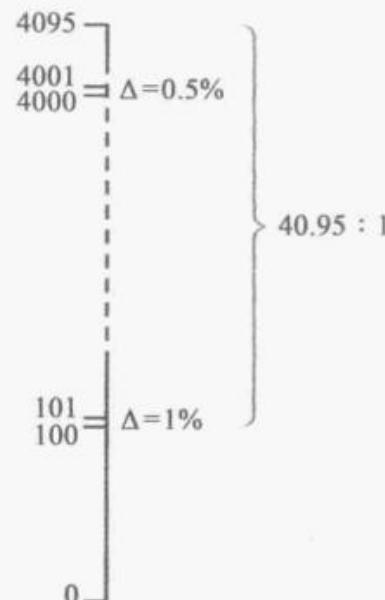


图 3-4 “编码 100” 问题的缓解

在视频（包括 motion—JPEG, MPEG 和 H. 264/AVC 标准）中，数字摄影领域（包括 JPEG/JFIF/Exif 标准）， $R'G'B'$ 分量采用均匀感知的方式编码。一个视频显示含有从码值到照度值的非线性转换函数，这个函数可以媲美绘制在图 1-7 上的功能，并且接近人类视觉明度感知的倒数。感知均匀性是通过采用对每一个场景估计感知的三刺激值的非线性转换函数来实现——伽马修正，伽马修正参数的选择——在默认情况下，采用手动或者自动的方式。因此预期的图像是在基准视觉环境的参考显示器上获得的。因此伽马修正整合了感知均匀性和图像渲染应该考虑的因素。

在计算中真彩色图像由每一个 8 位的 $R'G'B'$ 分量表示，我将会在后面章节做进一步的介绍。每一个分量是从 0（黑色）到 255（白色）。灰度和真彩色数据在计算中通常会被编码，以表现出接近于感知的均匀性。这个步骤与光强度不成正比，而是按照感知的间隔均匀隔开。所需步骤的数量依赖于显示系统、观看环境，以及感知的整体特性。

参考显示屏以及视频工作室的观看参数在发展史上不太规范或者是根本就不规范。在第 37 章，作者总结了一套切合实际的参数。

视频标准比如 BT. 709 和 SMPTE ST 274 之前上建立了标准的光电转换函数（OECF），犹如视频拥有的场景被称为图像状态，或者就好像照相机没有（校准）。实际上，视频是有效输出（显示）的，最重要的不是 OECF 而是 EOCF！非线性编码是第 27 章的中心话题，其中，也会介绍伽马修正，那一章将会讨论 OECF 和 EOCF。

如果视觉阈值严格按照跨越全音阶的 1% 关系表现，那么亮度编码可以按照对数编码，对于一个 100 : 1 的对比度，需要 462 个编码值，对应于 9 位编码值。

第 24 章将会展开关于明度和亮度建模的原因，在大多数数字成像系统里我们把视觉对明度的灵敏度建模为一个幂函数，而不是将其建模为一个对数函数。由显示器产生的红色、

绿色或者蓝色原色光的照度是与编码值成正比上升到接近于 2.4 次幂函数的。
$$\frac{\left(\frac{255}{255}\right)^{2.4}}{\left(\frac{254}{255}\right)^{2.4}} \approx 1.01$$

1.01 计算了线性光在最高的 8 位编码值（255）和下一个最低值（254）的比率，由此可知人类视觉的 1.01 比率要求得到满足。随着码值的减小，该比率变大。但是随着照度的减小，满足视觉的“JND”照度比率变大。2.4 次幂函数被证明对于感官要求匹配得非常好。

8 位成像系统通常被声明拥有的动态范围为 255 : 1 或者 256 : 1。这个声明源自于假设图像数据的编码值与光线线性相关。然而，假设一个 2.2 幂函数或者 2.4 幂函数的显示

器：对于一个理想的显示器，与编码 1 相关联的动态范围接近一百万比一，大部分的 8 位图像数据是根据感知来编码的，例如 sRGB 标准。在实际中，显示器和它的环境的物理参数是限制在这个动态范围里的。

几十年里，阴极射线管（CRT）是用于电视接收机和台式计算机的主要显示装置。令人惊讶的是，CRT 所表现出的 EOCF 非常接近视觉的明度灵敏度的倒数！视觉的非线性光照响应与 CRT 固有的幂函数相结合使得显示编值（从历史上看，电压值）表现出感知的均匀性，如图 3-5 和图 3-6（相反）所示。对于电子成像技术的前半个世纪，CRT 的特点是感知均匀性的基础。大多数现代电子设备（比如 LCD、PDP，以及 DLP）不具有像 CRT 固有的物理幂函数。然而，信号处理电路施加任何传递函数是必要的，以至于该装置的表现就好像它有一个从信号值到显示三刺激值的 2.2 幂函数或 2.4 幂函数。

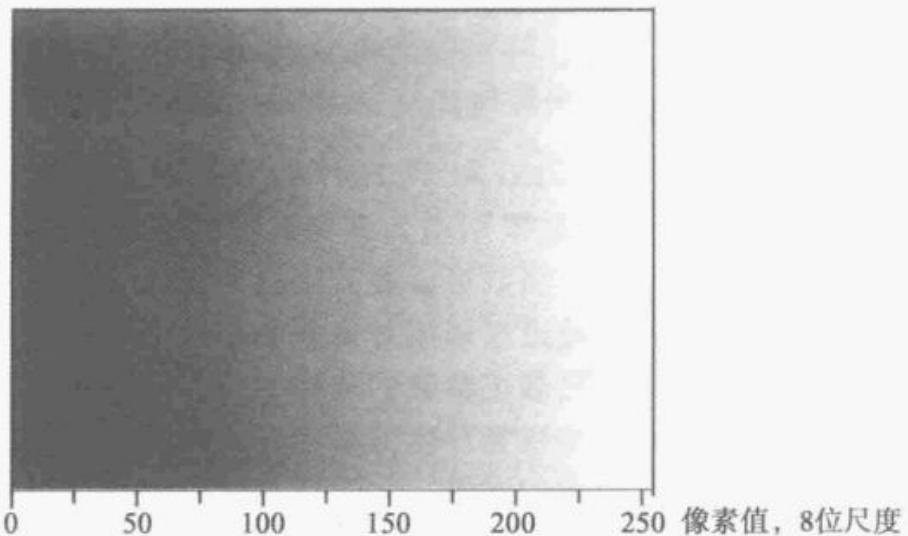


图 3-5 灰度渐变

灰度渐变在 CRT 显示器所显示的灰度坡道是通过把从 0 到 255 的连续整数写入帧的缓冲列生成的，经过了数/模转换器（DAC），呈递给 CRT 显示器，一个感知均匀性的明度扫描结果就出现了。一个天真的实验者可能会错误地得到一个结论——编码值正比于光强度

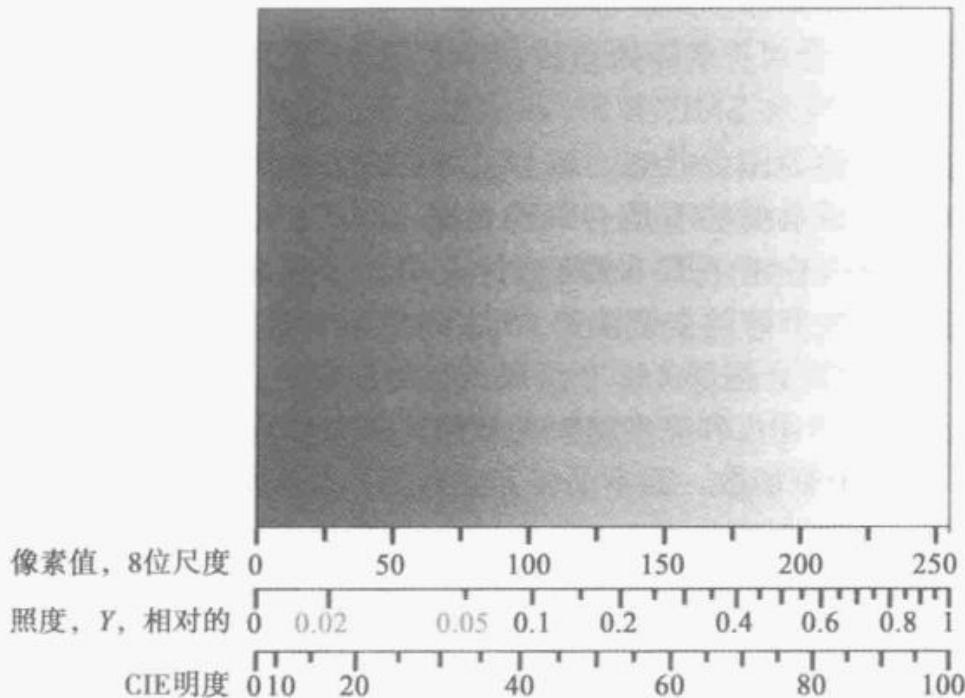


图 3-6 一个灰度坡道

由 CIE 相对照度增强 (Y , 正比于光强度，在量程的中间位置) 以及 CIE 明度 (L^* , 在量程的底部位置)。在屏幕上中间点位置的明度值是位于黑色和白色中间，编码值和明度之间接近线性关系，然而中间点的照度值大约是白色的 18%，由 CRT 产生的照度大约正比于码值的 2.4 幂函数。明度接近于视频信号照度值的 0.42 幂函数。令人惊讶的是这两者之间的关系近似互逆。这种近乎完美的消除，导致许多在视频、计算机图形、数字图像处理的工作人员误解了强度的含义，低估或者完全忽视非线性转换函数的重要性

在视频中，视觉上的感知均匀性被融入到每一个摄像机中的伽马修正电路使用。由伽马修正所得到的 $R'G'B'$ 值（这些值被处理），记录，并且发布到视频中——大概接近目标显示照度的 0.42 幂函数： $R'G'B'$ 近似是感知均匀的。感知均匀性允许用户用少到 8 位的值来代表每一个 $R'G'B'$ 分量。没有感知均匀性，每一个分量至少需要 11 位或者更多的数据来表示。数字静态相机图像编码标准，例如，JPEG/Exif 标准采用一个类似的方法。

3.5 线性和非线性

现代图像传感器如电荷耦合器件（CCD）和互补金属氧化物半导体（CMOS）有效地将光子转化为电子：它们产生的信号，幅度正比于强度，称为线性的。

以前视频信号是通过对电压具有线性响应的模拟电路来进行处理的。当今数字信号系统通常在像素值算法上是线性响应的。这种系统经常称为是线性的。

然而，如果非线性函数分开了两个领域，在一个领域的线性度就不能跨越到另一个领域。在视频中，场景照度在一个非线性光的领域，视频信号受电器信号领域的线性操作。然而，OECF 和 EOCF 被施加在这两个领域。因此照度和像素值不是线性相关的。当你问一个光学工程师，他的系统是否是线性的，他会说“当然！”——参考的是光的强度（intensity）、辐射度（radiance），以及照度（luminance）。当你问及视频工程师，他的系统是不是线性的，他会说，“当然是！”——他指的是像素值的算法。然而，在两个系统之间的是非线性转换：在一个领域的线性操作在另一个领域并不是线性的。

如果你的计算涉及感知，那么非线性表示是必需的。如果你在一个图像数据中执行一个离散余弦变换（DCT）作为图像压缩的一部分，例如 JPEG 标准的压缩图像，你使用的编码应表现出感知的均匀性，尽量减少由编码过程引入的感知误差。

第4章

量化

连续值幅度信号的量化是把幅度分配为若干（大约几百或者几千）间隔，使其成为一个个离散的数字化电平的过程。在均匀量化中，各个电平之间的步长具有相等的幅度。量化器丢弃了各个量化电平之间的信号信息。量化器的性能特征表现在这些信息损失的程度上。如图 4-1 所示是一个均匀量化器的传递函数。

4.1 线性度

电子系统通常被期望能满足叠加原理。也就是说，它们被期望能具有线性度，一个系统通常具有线性度，当且仅当它们具有如下的特征：

$$\begin{aligned} g(a \cdot x) &\equiv a \cdot g(x) & (\text{对于标量 } a) \\ g(x+y) &\equiv g(x)+g(y) \end{aligned} \quad (4-1)$$

函数 g 能包含一个完整系统：一个系统是线性的充分必要条件是，该系统的任何两个信号单独的响应的和等于两个信号和的响应。线性涉及稳态响应或者一个变化信号的时间响应。

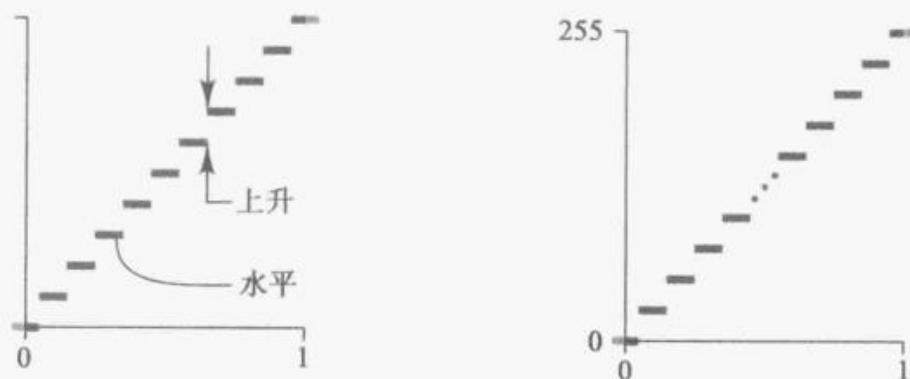


图 4-1 量化传递函数

线性在数学、信号处理，以及视频领域都是非常重要的特性。许多电子系统工作在线性度领域，并且使用信号直接代表物理特征。一个例子是，紧凑型音频光盘 (CD) 编码：声压级 (SPL)，与物理强度成正比，被线性量化到 16 位采样中。

人类的感知虽然是非线性的，在应用领域感知量被编码和传输，可以利用感知的非线性编码实现比原物理量编码更加高的效率。例如，数字电话的音频信号的非线性编码每个样本仅有 8 位 (A 律和 μ 律这两种在使用的编码，涉及的传递函数都可以与图 4-1 所示的双极版本相媲美)。图像信号的捕捉、记录、处理或者传送可以类似地采用非线性均匀感知的方式编码，以优化其性能。

4.2 分贝

在以下的章节中，我将会介绍信号幅度，噪声幅度，以及信号与噪声的比率——信噪比 (SNR)。在工程中，例如，SNR 通常是使用对数单位。 $10 : 1$ 的功率比定义为贝尔 (B)，目的是纪念 A. G 贝尔 (Alexander Graham Bell)。一个更加实用的比率是十分之一

的贝尔——分贝 (dB)，代表了功率比 $10^{0.1}$ ，大约是 1.259 倍。功率与功率比由式 (4-2) 给出。信号功率通常是相对于参考功率给出的，这个参考功率要么指定（通常单位是 dB），要么由式 (4-3) 给出。3dB 增加的功率值非常接近 2 倍 ($100.3 = 1.995$)，+10dB 的功率增加值与功率的 10 倍完全相同；-10dB 的功率减少值为功率的十分之一。

功率比使用分贝表示为：

$$m = 10 \lg \frac{P_1}{P_2} \quad (4-2)$$

功率比相对于基准功率为：

$$m = 10 \lg \frac{P_1}{P_{REF}} \quad (4-3)$$

功率比以分贝为单位，作为电压的一个函数为：

$$m = 20 \lg \frac{V_1}{V_2} \quad (4-4)$$

设想一个电缆传输一个 100MHz 的无线电频率信号。经过 100m 的电缆，功率已经减小到一定的分数，可能是原来信号的 1/8。再经过 100m，会减小到原来的同样的比率。我们不是表达为该电缆每 100m 衰减为原来的 0.125 倍，而是表示为每 100m 衰减 9dB；1000m 电缆的末端功率值为参照的源功率的 -90dB。

分贝定义为功率比，如果施加一个恒定电压源，电压加倍，电流也会加倍，所以功率增加了 4 倍。更一般地讲，如果电压（或电流）加到一个恒定的阻抗上，以比值 r 变化，功率的变化采用 r^2 ，($\lg r^2$ 等于 $2\lg r$)，使用式 (4-4) 从一个电压比计算出分贝。在数字信号处理 (DSP) 中，数字编码电平被等效地处理为电压，在数字信号处理中的分贝是基于电压的比率得出的。历史上模拟信号系统中普遍使用的一个参考是 1mV (dB_{mV})；在数字系统中，参考通常是从参考黑色到参考白色 (dB_{FS}) 的全尺度变化，相当于 8 位接口电平的 219 编码。注意：历史上的 8 位计算机图形处理 8 位信号没有下余量也没有上余量。实践中在 MPEG 领域发现 PSNR 算法，在 MPEG 领域有一个满量程由 0~255 来取代 0~219。

表 4-1 给出了用分贝来描述电压比的例子。

频率比为 2 : 1 是一个关于音乐的八度值，指的是涵盖 2 : 1 频率范围中音乐的 8 个整乐：do、re、me、fa、sol、la、ti、do。电压减半，频率加倍，电子工程师称此为每八度音 6dB 的损耗。如果频率每翻一番，电压减半，电压减到十分之一，频率增加到 10 倍：10 : 1 的比率是一个十进制的量，所以 6dB 每八度音等于 20dB/10 倍频程（以 2 为底 10 的对数非常接近 20/6）。

在摄影中一个停止的比例是 2 : 1 的光功率比，正如上面提到的，1 分贝为功率比 $10^{0.1}$ ，或者 1.259。传感器和摄像工程师更喜欢使用在光学和电子领域等效的单位：它们将数字编码级别当做信号（就像电压），描述 2 倍的光功率为 6dB。 $10^{0.3}$ 非常接近 2，这是一个数字巧合。所以 6dB 对应一个停止位，2dB 对应 1/3 个停止位。

表 4-1 分贝举例

电压比	分贝
10	20dB
2	6dB
1.112	1dB
1.0116	0.1dB
1	0dB
0.5	-6dB
0.1	-20dB
0.01	-40dB
0.001	-60dB

4.3 噪声、信号和灵敏度

模拟电子系统不可避免地会引入来自热源或者其他源的噪声。热噪声与被处理的信号并

不相关。系统也可能会受到外部干扰源的干扰。当信号幅度减小时，噪声和干扰影响较大。

处理、记录和传输都会引入与信号不相关的噪声。此外，与信号相关的失真也可能被引入。由于其涉及关于系统性能的客观测量，失真将会视作是噪声，但是给定量的失真相比于等量的噪声可以被或多或少的感知。失真可以归因于一个特定的处理过程，称为伪影，尤其是，当它有一个独特的视觉效果的时候。

在视频中，信噪比（SNR）是一个特定信号幅度的峰峰值，通常为系统承载的基准振幅或者最大幅度除以无用分量的均方根（RMS）来表示，无用分量包括噪声和失真（对于一些强调峰值的信号，有时候称为PSNR，见图4-2），信噪比以分贝为单位。在很多领域，比如，音频，信噪比被指定或者是在物理（强度）领域测量。在视频中，信噪比通常被应用于伽马修正的分量 R' 、 G' 、 B' ，或者在感知领域的 Y' ，所以信噪比与感知性能有关系。

灵敏度是指可以接收（或者指定）信噪比的最小源功率。

4.4 量化误差

下式为 k 阶量化的SNR理论限制：

$$\text{SNR} = 20 \lg (k \sqrt{12}) \quad (4-5)$$

量化信号只携带离散的，预定的电平：相对于原来的连续信号，量化误差已经被引入，这个误差与信号有关，并且可能会称为失真。然而传统的信号理论用来处理作用于信号的噪声。相对于信号的幅度每个量化的步长假设较小，我们可以把量化增加的误差作为增加了等量的噪声：量化减小了信噪比。式(4-5)给出了 k 步长量化理论信噪比限制。在消费类视频仪器中8位量化有一个大约56dB的理论信噪比限制（峰峰值除以方均根（RMS）噪声）。

如果一个模拟信号的噪声非常小，当接近一个步长时它的量化误差几乎可以确定，当模拟信号在量化电平之间时，差不多可以表现出接近 $1/2$ 个步长的误差。在视频中，这种情况可以导致重现的图像会显示出噪声调制。在量化之前引进大约有二分之一量化步长的高频随机噪声或者伪随机噪声去避免这种作用是很有好处的。这就引入了一个小的噪声进入画面，可是这种噪声与低频量化“图案”相比不太明显，如果没有这种噪声，低频量化噪声会比较明显。信噪比略为减小，但是主视图的质量却大大提高。历史上，视频的数字化隐含着假设输入信号具有足够的模拟噪声去实现其功能。如今，模拟噪声水平较低，应该明确针对数字化端增加噪声。

视频信号中的噪声可见的程度——或令人反感的——依靠于视觉特性。为了减少噪声的可视度，我们一个信号采用精心选择的非线性照度（或者三刺激值）函数数字化一个信号。这个函数被选择，以至于从黑色到白色整个色调范围内，一定数量的噪声能近似等价感知。这个概念在非线性图像编码有相应的概述。在接下来的章节中，线性度和感知均匀性将会被详细阐述。

4.5 全摆幅

漂移（或者通俗上讲的，摆动）是指一个信号的范围——最大最小值电平之间的差值。在视频中，参考偏移是标准的参考白色到参考黑色信号电平之间的变化值。

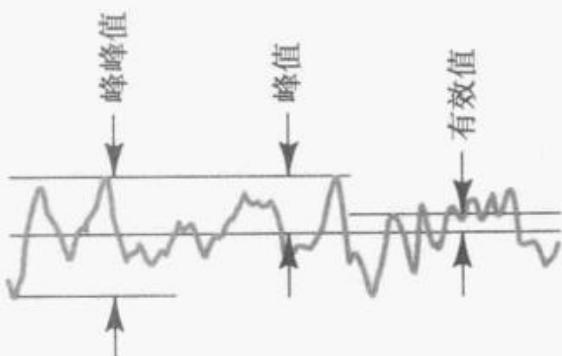


图4-2 峰峰值、峰值、有效值

通常计算机图形数据是全摆幅的（或者说是满量程的），参考黑色是最小的编码电平，参考白色是最大的编码电平。在桌面图形中， $R'G'B'$ 分量通常是由 0~255 变化的 8 位组成。图 4-3 描述了 8 位全摆幅编码。

4.6 演播室摆幅（下余量和上余量）

在高品质的视频中，有必要保持瞬态信号偏移低于黑色和高于白色，这经常是经过数字滤波器和模拟滤波器处理后的结果。因此，演播室视频标准要求下余量低于基准黑色，上余量高于基准白色。由于上余量允许编码值超过参考白色，参考白色和峰值白色是指的不同的信号水平。

在数字视频标准里提供了下余量和上余量。进行处理时参考黑色被认为是在代码零处的一个欺骗；在 8 位系统中， $R'、G'、B'$ 以及亮度符号 (Y') 从零到参考白色编码值为 219 的范围内变化。在一个 8 位接口中根据 BT. 601 标准中添加了 +16 的偏移量（以斜体字表示）。接口编码值 0 和 255 被保留，用于同步处理；这些代码值禁止出现在视频数据中。

本书已经在抽象系统领域描述了 0 到 1 的变化。当 $R'G'B'$ 或者 Y' 参量由 8 位作接口，从 0 到 1 的值就缩放 219，偏移是 16。8 位演播室标准因此就有了从参考黑色到参考白色的 219 个步长，接口编码值低于 16 和高于 235 的用来做下余量和上余量。不幸的是，下余量和上余量是不对称的。图 4-4 展示了 8 位 $R'G'B'$ 或者亮度的标准编码范围。具有 8 个最重要位全 0 或者全 1 的编码，0 到 255 用于同步处理；这些代码禁止用于视频数据。

历史上，有两个下余量和上余量存在的原因。

- 1) 对于适应模拟电平，被误调的视频是有必要的。
- 2) 对于通过尽可能多的系统（尽管下冲最终会在显示器中被裁剪）时，保持过滤瞬变（下冲和过冲）是很重要的。

由于模拟使用的减少，第一个原因消失了。第二个原因仍然很重要，另外三个原因出现了。

3) 下余量可以使能 PLUGE 模式的传送用于设置黑电平。

4) 当黑色被正确地设置，参考黑色电平产生的照度在视觉上难以与基准黑色区分开来，但仍然是非零的。一个纯粹的 2.4 幂电光转换函数 (EOCF) 产生的相对照度大约在 0.0003 左右。理想化地，理论为零的照度出现在低于基准黑色电平处。该电平编码在下余量区域。

5) 图像传感器产生的噪声包括黑色区域的噪声。如果提供了下余量区域，噪声的平均值就为黑色。如果缺少了下余量区域，就会使噪声向负向偏移，噪声的平均值就会上升。提供的足下区域阻止了噪声调制。

因此，第一个原因过时了，其他的四个原因仍然是很重要的。

至于上余量，参考黑色最初被建立的电平稍微低于峰值白电平，为的是适应模拟调整

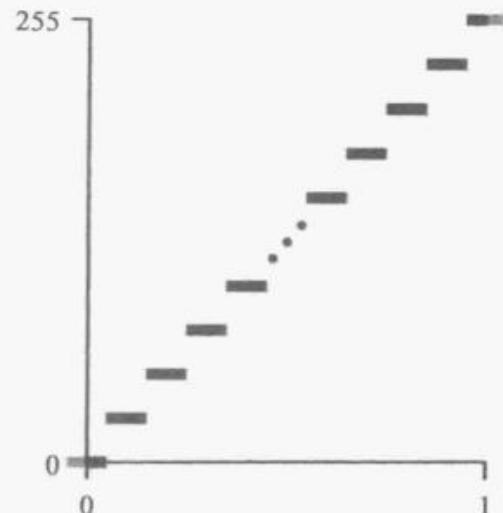


图 4-3 全摆幅的 8 位量化

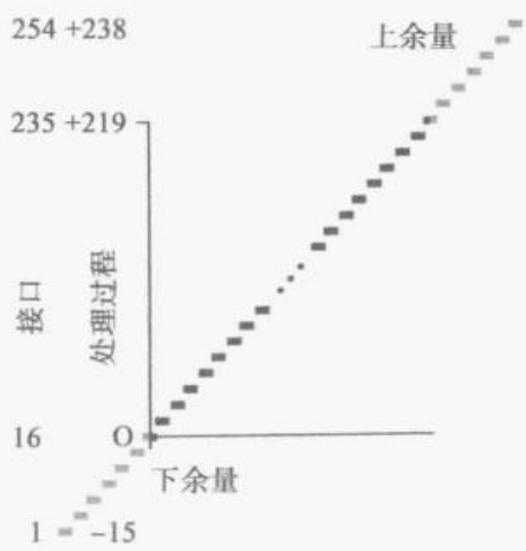


图 4-4 下余量和上余量

的失误。随着数字化的盛行，那个原因消失了。但是参考白色对于数字视频的地位更加稳固，还配有 2.4 幂值的电光转换函数（EOCF），峰值白色的相对亮度是 $(955/876)^{2.4}$ 或者大约 1.23。演播室的参考显示一路到峰值白色应该遵循 2.4 幂函数曲线。

消费类电子总是致力于最大限度的提高平均明亮度，并且这个目标受到在高端平坦化电光转换函数（EOCF）的常规辅助。在节目制作时，假设 CE 设备可以合理可靠地显示电平直至参考白色电平，然后欺骗。

在具有 k 位 ($8 \leq k$) 的接口中，参考黑色电平和白色电平乘以 2^{k-8} 。例如，当 $R'G'B'$ 或 Y' 的分量占据 10 位，0 到 1 的范围被扩展 219×4 （也就是从 0 到 876），则接口的偏移量是 +64 而不是 +16。

超级黑色和超级白色这两个术语在消费类电子领域使用时其含义分别是指进入上余量和下余量。在一些消费类的装置（例如 PS3）中，超级白色是一个选项，此时可以提供一个接口去执行下余量区域和上余量区域的编码。当不设置的时候，接口编码就会被剪切，摄影室标准材料的质量就容易受到影响。

我用演播室摆幅这一名词代表演播室设备所用的电平；使用全摆幅这一名词代表在接口极限处的参考水平所用的编码，这个极限水平没有上余量和下余量。在历史上看来，桌面图形编码有全摆幅（从 0 到 255）。一个实时的 HD-SDI 接口禁止从 0 到 255 编码或者在 10 位的系统中禁止 0~3 以及 1020~1023 的编码。一个全摆幅编码是跨越一个 HD-SDI 接口编码的，必须避免 0~3 以及 1020~1023 的代码。

在数字影院的采样以及后期制作中，通常会在下余量区域保留底部的 64 位码，但是使用的所有的上余量可以达到 1019 的码值。一些数字影院系统称这个为“扩展”范围，但是在我看来，使用“扩展”这个词会让人误以为使用这个范围是一个好主意（谁不想要“扩展”呢）。在我看来，这个词语应该用“极端”代替。

4.7 接口偏移量

在硬件中，一个 8 位的接口可以认为用来传递 0 到 255 的值。在串行数字视频接口（SDI，或者它的 HD 变量，HD-SDI）中，所有的 8 位的全 0 和全 1 数字编码禁止使用。在一个 8 位数字视频接口中，编码值添加了 +16 的偏移，如图 4.4 的黑体所示：参考黑色的编码是 16，参考白色编码是 235。我认为应在接口处增加或者删除了偏移量，因为对于许多处理操作（例如改变增益）有符号的表示法是必要的。然而，硬件设计人员通常考虑 8 位数字视频的黑色代码是 16，白色代码是 235。使用 16 代表黑色会让界面的设计变得简单，但是会使信号的算法设计变得更难。

4.8 处理编码

在信号处理中，使用独立于编码范围的 0 代表参考黑色是很方便的（有的时候是必要的）。为了适应下余量，数字表示必须使用负数。在抽象层面上描述信号处理——或者在浮点运算中实现信号的处理——使用范围 0~1 是最简单的方式。这个参考是把 0 和 1 点分别作为参考黑色和参考白色（这个范围也称为单元，在这个范围内从参考黑色到参考白色有 100 个单元）。为了在上余量适应信号，这个数字代表必须允许比单元更大的数字，在下余量也需要小于零的值。

在硬件处理时，样本通常用来代表一个具有有限位数字的固定点整数。使用二进制补码算数通常是很方便的。信号处理所需要的位深度通常比接口处需要的要更大。黑色通常编码为 0。参考白色将被编码为适当的值，例如在 8 位系统中的 219 或者在 10 位系统中

的 876。

在信号处理中，即使是没有接口偏移，处理负数也是很必要的。二进制补码表示法很常见。

$R'G'B'$ 或者 $Y'C_B C_R$ 的每一个 8 位的分量足以满足消费类视频的分发。但是，如果一个信号必须被处理多次，比如说，纳入一个多层复合图像，舍入误差易于积累。为了避免舍入误差，演播室的数据通常每一个 $Y'C_B C_R$ 分量有 10 位数据。10 位演播室接口有 4 倍的图 4-4 和图 4-5 所示的参考电平：有 2 位额外的数据被追加为最低位，来提供更高的精准度。在设备处理数据时，中间的结果可能被保持为 12、14 位，甚至是 16 位。

图 4-4 展示了一个单极信号，例如亮度的量化器。 C_B 和 C_R 信号是双极性的，在正值和负值之间变化。对于 C_B 和 C_R 的标准是使用一个水平中点型量化器，例如在图 4-5 所示的绘图中，从而使零色度具有一个准确的表达。在处理过程中，有符号的表示是必要的；在演播室视频的接口中，标准上是扩展 8 位色差分量到 224 的偏移，并且添加一个 +128 的偏置（请注意色度比亮度多占据 5 个 8 位编码）。

4.9 二进制补码回绕

现代计算机使用二进制数表示。有符号的整数运算是使用二进制补码表示的。当算数运算的结果例如加法或者减法溢出固定的可用位深度时，二进制补码运算回绕。例如，在 16 位的二进制补码运算中，取最大的正数，32767（用十六进制数表示是 $(7fff)_h$ ）添加 1，产生了最小的负数 -32768（十六进制数表示是 $(8000)_h$ ）。这是一个使用计算机软件实现视频算法的隐秘的难题，视频算法中整数运算允许使用回绕。在视频信号处理中，必须防止回绕而且还要使用饱和算法。

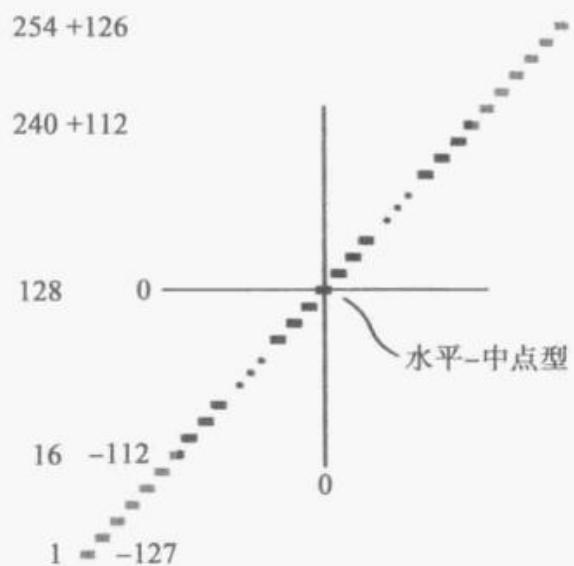


图 4-5 水平-中点型量化器

C_B 和 C_R 双极性信号水平-中点型量化器允许精确地表示零色度（在视频中，上升-中点型量化器很少使用），在处理中， C_B 和 C_R 抽象的取值范围是 ± 112 。根据 BT.601 标准，在 8 位演播室视频接口中，加入了一个由斜体值表示 +128 偏移。像表示亮度一样，接口处编码值 0 到 255 保留用于同步处理。

第 5 章

对比度和明亮度

几乎在所有的电子显示屏上都可以找到标志着对比度和明亮度的用户访问控制标记。这些标记连接到明亮度和对比度的感知属性不是直接的而且容易混淆。在 CRT 中，从最佳设置向上调整明亮度比可比较的对比度控制调整更加影响视觉对比度。调整对比度相对于可比较的明亮度控制对视觉明亮度的影响更强。因此，明亮度和对比度是误导性的标签。现今，明亮度和对比度的控制是在数十亿件的设备中实现的。近半个世纪，数亿人对这些控件缺乏了解：成像系统的设计人员面临一个很大问题。

本章讲述明亮度和对比度的感知属性、传统的对比度和明亮度的控制，解释了这些控件是如何做它们应该做的事情的，在最后本章还提出了一些建议来减少这个疑惑。

5.1 感知特性

根据两名备受尊敬的视觉和显示系统的研究专家的描述，至少对非专业的观众来说，当评估高端电视的图像质量时，有 4 个最重要的图像质量属性，具体是明亮度、对比度、色彩渲染和清晰度。

下面着重介绍两个图像属性：明亮度和对比度，这也是作者认为是最重要的两个属性。

Heynderickx 和她的同事指的是明亮度和对比度的感知属性。在一些显示设备中有类似命名的控制器，但是作者认为，这些控制不是用一种比较明显的方式影响显示图像感知属性的。在本章，包括标题，我们必须从感知属性上区分控制器的名字。本章关于控制器的名字使用楷体表示——对比度和明亮度——视觉属性中的明亮度和对比度字体未特别标注。

对比度是指颜色或者灰度之间的测量或者视觉区别。对比度通常指高照度值（或者反射率）与低照度值（或者反射率）的比率的。比率可以在大范围的不同照度值之间进行计算。例如，当评估一个显示系统时，我们通常选择 100 或者更好的对比度（最大照度与最小照度的比值），有的时候这个对比度会达到 10000。这个比率可以在类似的照度之间进行计算。当比率小于 1.01（“韦伯定律”）时，视力不能区分两个照度水平。接近人眼区分阈值的两个照度的比率有时候称为韦伯对比度。

5.2 显示处理信号的历史

电视起源于模拟电子管电路；CRT 显示器本身就是真空管。真空管和相关的模拟元件（主要是电阻和电容）由于工作温度的变化以及长时间使用引起老化，容易产生漂移。漂移的主要影响是改变增益和视频信号的偏移量；因此，增益和偏移需要控制。

用户可调节对比度和明亮度的控制器，在 20 世纪 40 年代初的电子管电视接收机中已经实现。视频放大器电路的增益由后来称为对比度的控制器所调整。偏移（偏差）的控制由 CRT 本身实施，使用一个叫做明亮度的控制器来完成。增益控制比偏移在信号路径中生效早。Kallmann 描述了一种典型的实施：

……所谓的对比度……是一个分压器控制信号幅度……背景光的控制……调整阴极射线管的偏压。

Kallmann 描述的方案占了整个 CRT 时代的上风。对比度和明亮度电路工作在 $R'G'B'$ 领域——那就是工作在伽马编码信号。从历史上看，在 CRT 本身施加的幂函数与显示器伽马（Gamma）曲线有关。在 CRT 显示器中，伽马是不可调的。

我还一直无法找到任何历史文献来介绍对比度和亮度这两个名字是如何来的。一些早期的电视接收器使用光彩标签标识增益控制；一些使用背景做偏移控制；一些早期的电视型号有同心圆形对比度和音量控制器，表明这个位置是让用户用来改变声音的大小和图像大小的；一个模型是在前面板的垂直保持和焦点之间有亮度控制。

视频科学家、工程师和技术人员质疑明亮度和对比度的名字几十年了。60 年前，Oliver 写道：

增益（“对比度”）控制肯定会比偏置（“明亮度”）控制产生更多的接近纯粹明亮度变化。因此在一定意义上，按钮是贴错了标签。

括号和引号是原有的。对于明亮度，Oliver 说：

旋钮比较好听的名字是“黑色”、“背景”或者“阴影”。

一个卓越的电子工程师 Donald Fink 通过观察几年之后得出，这些控件都名不副实。

“不幸的是，在现如今的电视系统中……接收机的明亮度和对比度的独立操纵（对于光度学而言，这两者都是名不副实的）被不懂技术的观众使用之后，可能会随时使得系统设计人员和操作技术人员所做的最佳努力结果付诸东流。”

在我看来，不幸的是，尽管几十年来像 Oliver 和 Fink 研究学者这么抱怨，这些名称仍旧使用。

70 多年来，视频处理技术不断发展，1965 年第一次出现在模拟模式下使用的晶体管，接着就是大约 1975 年的模拟集成电路，然后在 1985 年的数字集成电路中，它的复杂性在这 25 年中大大提高。在 2000 年左右，显示技术开始从 CRT 向 LCD 和 PDP 技术转换。经过了这些变化，需要进行的调整变小了。尽管如此，对比度和明亮度已经进一步发展（不用多想，有些人会说）为技术的连续几代的模式。如今，这些控件使用在 10 亿的 CRT 电视接收机和另外 10 亿的计算机 CRT 显示器中。这些控件已经发展成为固定像素的显示（无需多想）；围绕如今 10 亿的 LCD 显示屏，几乎在所有数字信号处理路径中都有明亮度和对比度控制器。

在视频处理设备中，增益和偏移控制器在历史上都是可以使用的；它们与显示控件同等操作，但是相关的控件通常标有增益和黑电平。

液晶和等离子显示器通常具有明亮度和对比度控制器。尽管专业用户期望控制器可以同样执行如 CRT 显示器类似控件一样，尽管消费者期望这些控件应该使用类似的方式在 CRT 中运作，液晶控制器通常具有不同的效果。

对比度和明亮度控制广泛存在于计算机的图像应用中。在这些领域中对比度和明亮度控制不是必须与显示设备中类似控件名字的效果相比拟。特别是，在 Photoshop 中对比度表现与典型显示设备的不同：Photoshop 对比度控制着增益，但是它的“扭转”是围绕着一个平均像素水平的特定公式扭转，而不是像通常在视频设备和显示设备中那样围绕零值进行扭转。对于其他计算机图像应用，没有标准；往往很难或者是不可能准确地表明一个特定的应用程序如何实现这些控制。

5.3 数字驱动电平

数字驱动电平（DDL）这个术语是指通过一个接口（通常是 DVI，HDMI 或者是显示端口）和驱动显示设备的像素分量数据值（典型的是一台 PC 图形子系统或者是消费电子信号源比如蓝光播放器产生）。一个 DDL 被解释为一个 $0 \sim 2^k - 1$ （其中 k 是接口的位

深度，典型值是 8，也可能是 10 或者 12）的整数值。

计算机接口，比如 DVI，携带 8 位 DDL，其中的 DDL 0 值表示参考黑色，DDL255 值表示参考白色。

视频接口标准比如工作室里的 HD-SDI 以及消费类电子设备 HDMI 都允许下余量在参考黑色以下，上余量在参考白色以上。HD-SDI 是标准的 10 位值，接口编码 64 表示参考黑色，接口编码 940 表示参考白色。消费者经常使用 8 位的 HDMI；接口编码 16 表示参考黑色，接口编码 235 表示参考白色。

为了简化讨论，本文将会参考使用归一化 DDL（NDDL）的像素值，其中参考黑色在接口处对应于 NDDL0，参考白色对应于 NDDL1。允许视频中的像素值在参考 0~1 范围之外适度偏移，范围大概是 $-15/219 \sim 235/219$ （大约是 $-0.07 \sim +1.09$ ）。NDDL 的 0~1 的变化范围对应于 HD 工程师称为 IRE 的 0~100 的电平范围。

5.4 信号和明度之间的关系

sRGB 标准要求 EOCF 包括图 5-1 所示的 2.2-幕函数。

与 sRGB 标准相关的文件明确指出基本的 EOCF——从像素值到显示照度的映射——应该是一个纯粹的 2.2-幕函数添加了一些杂光的术语。sRGB 标准还会记录旨在描述从显示照度到像素值的映射函数 OECF，这个函数适用于模仿摄像机，在那里黑色处反幕函数的无限长斜率将是一个难题。该 OECF 有一个接近黑色的线性段，还有一个指数为 1/2.4 的幕函数段。这个 OECF 不应该被反转用作 EOCF，因为近黑色直线斜率不适用于 EOCF。

sRGB 标准的 2.2-幕函数，实际上现在演播室参考显示器的 2.4-幕函数，以及数字电影院的 2.6-幕函数都几乎完美地反转了 L^* ，如图 5-1 所示。2.4-幕函数曲线，代表了视频和实际中的高清，与代表 NDDL 0.2 或者更高版本（也就是大于 59 的 8 位接口编码）的 L^* 具有高度线性关系。一个值为 0.2 的 NDDL 产生一个值为 16 的 L^* 。从（0.2, 16）点通过参考白色的直线具有 105 的斜率；向后延伸那条线在 x 轴产生一条截距为 0.0475 的 x 值，更进一步向后， y 截距大约是 -5。

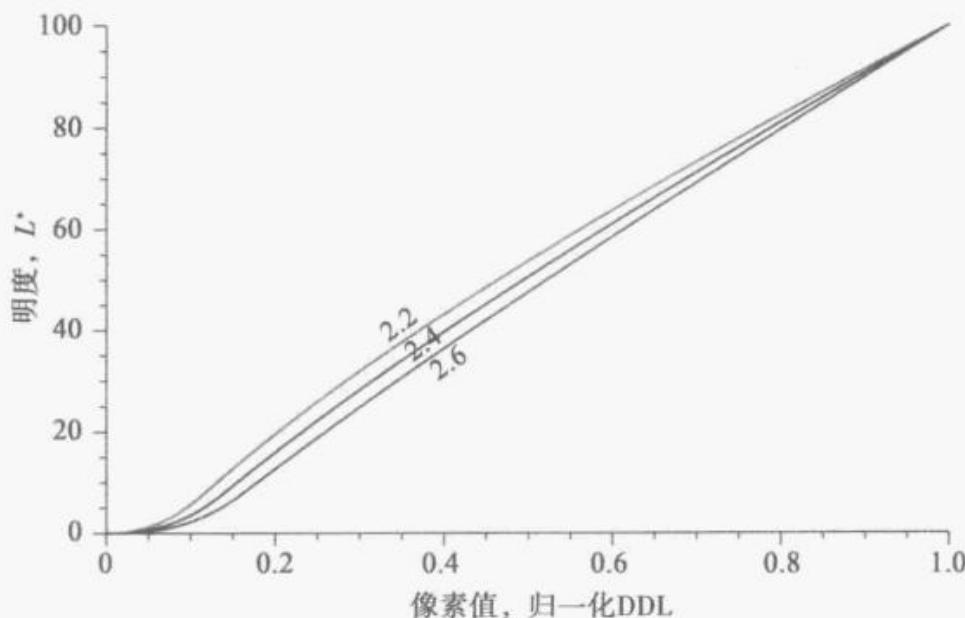


图 5-1 像素值和 L^* 的关系

在 x 轴绘制归一化像素（或者视频信号）值。相应的 L^* 值绘制为 2.2-幕函数（典型的计算机图像，例如 sRGB 标准），2.4-幕函数（典型的演播室视频），2.6-幕函数（标准的数字电影）的显示幕函数（“伽马”幕函数）。在该图中，以及接下来的，当 $R=G=B$ 的适用于三个信道，或者适当的缩放（其中，另外两个信道为 0）的单个通道

5.5 算法

在视频中传统的对比度和亮度控制效果是由下面的公式近似， x 代表了范围为 0~1 的 NDDL (R' 、 G' 或者 B' 之一)：

$$y = mx + b \quad (5-1)$$

图 5-2 所示的是对标称偏移量（“辉度”）设置 ($b=0$) 控制增益的影响（“对比度”），这展示了 0.8、1.0 以及 1.25 的增益值，这些都是在式 (5-1) 中 m 的值。 y 轴是三刺激值，不是直接感知含义上的线性光亮。

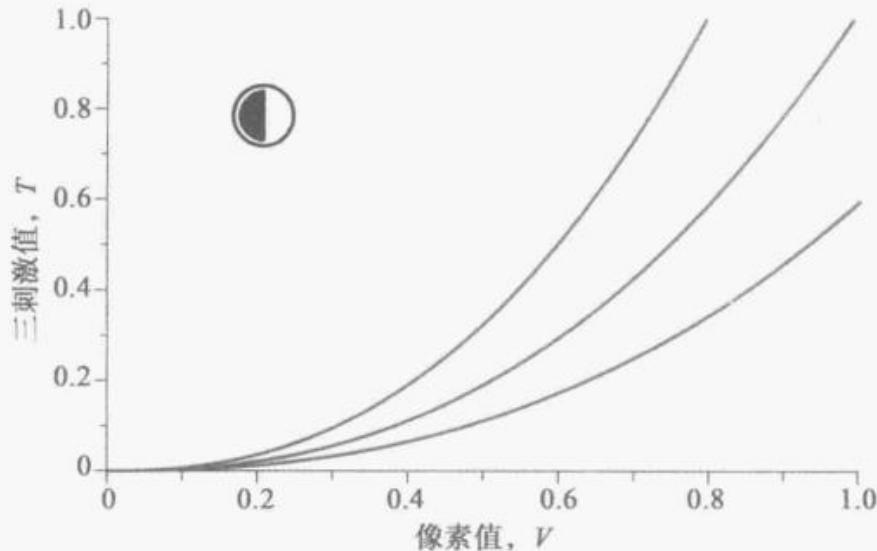


图 5-2 控制增益的影响

图 5-3 所示的是对标称增益（对比度）设置的偏移控制（明亮度）影响。展示了一 0.2、1.0 以及 +0.2 的偏移值，这些值都是在式 (5-1) 中 b 的值。

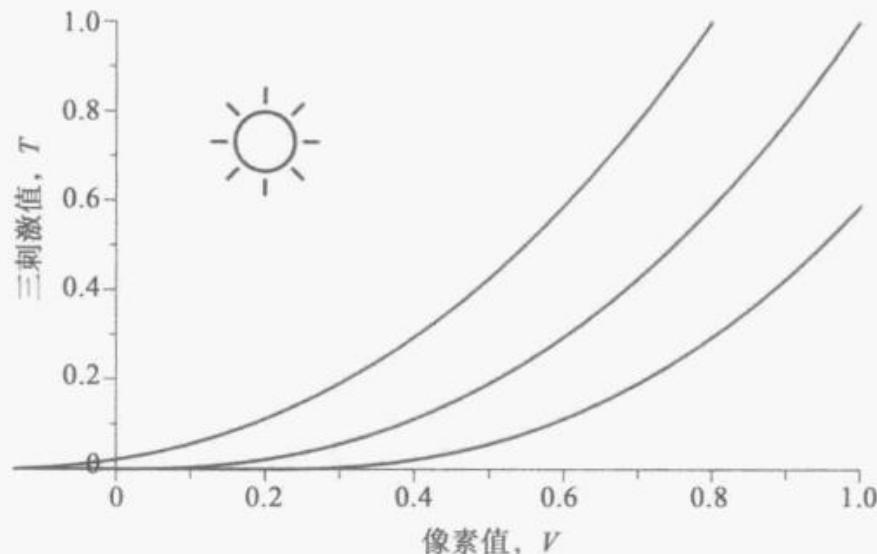


图 5-3 偏移控制的影响

对比度改变 m 参数——Oliver 和 Fink 称它为增益——在一个接近 0.5~2 的变化范围。明亮度改变了 b 参数——Oliver 和 Fink 称它为偏置或者偏差——变化范围接近 ± 0.2 。

图 5-2 和图 5-3 分别展现了改变增益和偏置的三刺激值显示效果。

式 (5-1) 的结果被 y 值截断（以前是 CRT 本身的作用，现在是通过信号处理），所以没有零以下的值。在不会超过 1.09 太多的足够高的信号值中，饱和度可以被设置。用户在调节增益和偏置时应当注意自己的设置不要引入截断或者饱和度。

式 (5-1) 的 x 和 y 信号在伽马修正 ($R'G'B'$) 领域。然后结果被提升到一个温和的

功率 γ (伽马值, 变化范围是 2.0 到 2.6), 进而产生一个显示三刺激值 (R 、 G 或者 B , 在图 5-2 或者图 5-3 的 y 轴)。

历史上模拟增益控制电路把对比度的实现为钳位在消隐电平 (0) 上的 $R'G'B'$ 视频信号的“第 1 象限”的乘数。

在 PAL 视频中, 黑色和消隐电平分别为相同的 0 值; 因此, 调整 PAL 接收机的对比度时留下了一个正确编码信号, 它被认为是黑色的 (具有的亮度交互最少)。在 NTSC 编码中, 插入了 +7.5 单位的“设置”, 导致了一个正确编码的黑色电平位于 $0 \sim 1$ 范围的 0.075 的位置。然而, 钳位仍然位于消隐电平 (0)。因为在零位“铰接”, 从 $0.5 \sim 2$ 调整增益会转换黑色电平变成 $0.0375 \sim 0.15$ 的范围, 由此导致了对比度和亮度进行了交互。

m 和 b 变化范围, m 和 b 值与控件的关系, 或者呈现给用户的数字控制值, 都没有相应标准或者约定。当今的演播室的参考显示器 (“BVM”) 是可调的, 允许 m 的变化范围是 0.5 到 2 之间, b 的变化范围是 ± 0.2 ; 然而, 在演播室的实际应用中, 两个控件都缺乏数控值是非常常见的。在当今的消费设备中, 呈现给消费者的典型对比度是一个从 0 到 100 的值 (这里标志为 C) ——作为屏幕显示器 (OSD) 的一个实例——亮度是一个从 0 到 100 变化的值 (这里表示为 B)。从控制值到参数 m , 以及 b 的合适映射如下, 其中 m 从 $0.5 \sim 2$ 变化, b 有 ± 0.2 的变化:

$$m = 2^{\frac{C-50}{50}}; b = \frac{B - 50}{250} \quad (5-2)$$

亮度可以交替呈现的范围是 $-50 \sim +50$, 在这种情况下, 第二映射将是 $b = B/250$ 。

图 5-4 描述了通过考虑参数方程的显示特性, 针对增益参数 m 的不同设置而产生的明度。图中值的范围对应于在式 (5-2) 映射下对比度 $0 \sim 100$ 的变化范围。

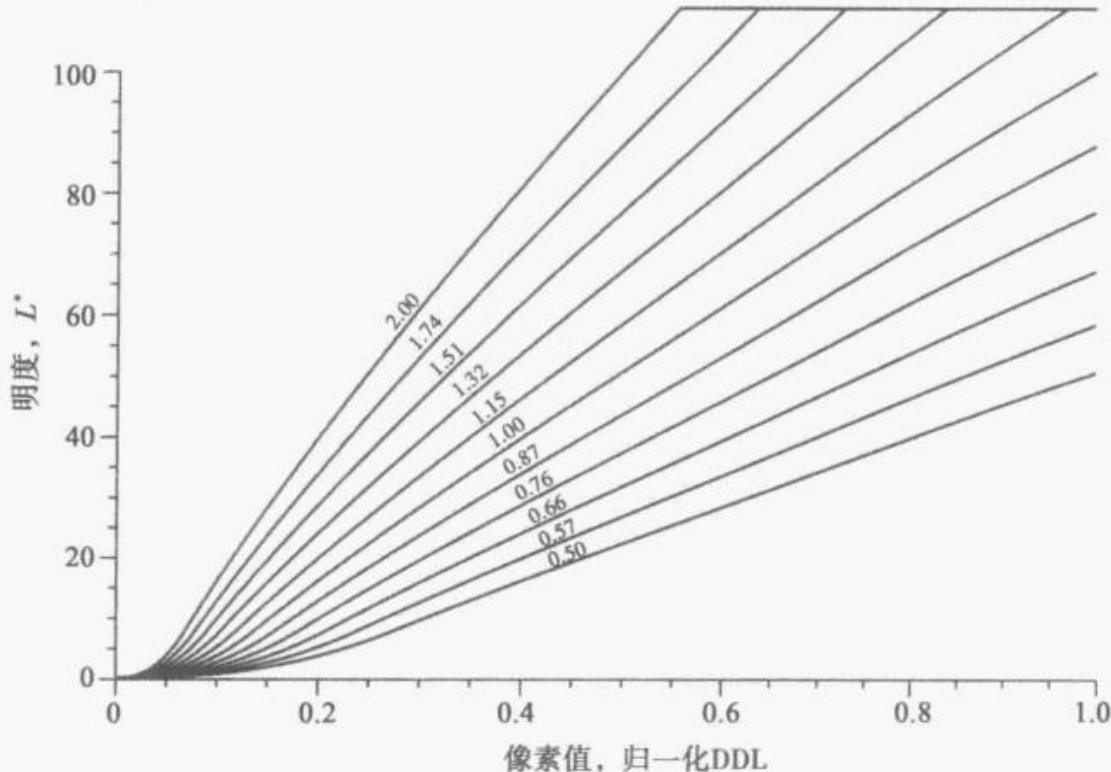


图 5-4 增益控制效果

图 5-5 展示了通过考虑该参数函数的显示特性, 针对偏移参数 b 的设置而产生的明度。图中值的变化范围对应于在式 (5-2) 的映射下亮度的 $0 \sim 100$ 的范围。

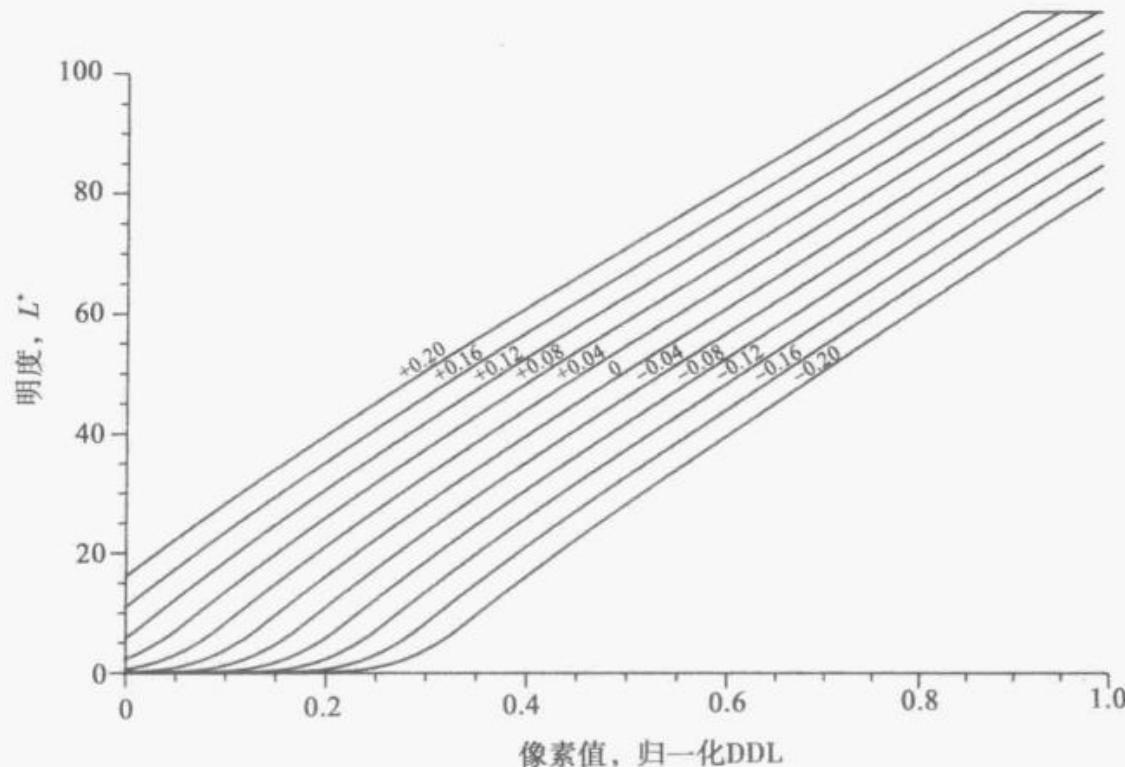


图 5-5 偏移控制效果

5.6 黑色电平设置

要在工作室设置明亮度（或者黑电平），需要在一个具有平均相对照度约 0.01（1%）的测试图像显示一个含有 PLUGE（级别为 $-0.02, 0, +0.02$ ）的模式。设置黑色高电平，然后减小，直到 -0.02 和 0 的 PLUGE 模式变得难以区分为止，明度设置就完成了。

如果你没有 PLUGE 模式，则显示一个大部分是或者主要是黑色的图像。把黑色电平设置到最低，然后提高电平，直到显示器几乎不再显示深灰色的提示为止，然后再下降一点电平。

一般来看，黑电平的设置有一点依赖于环境的光亮。然而，现代的显示器具有如此低的面板反射率，以至于环境产生的不需要的照度非常小，黑电平的设置不再对照明环境敏感。现代显示设备非常稳定；没有必要频繁地调整。

在结束时，8 位编码 $0 \sim 16$ 可能没有办法区分。代码 16 (NDDL 0) 被期望产生的照度从 PLUGE (8 位接口编码 12, NDDL -0.02) 的负向在视觉上可以区分：PLUGE 正方向 (8 位接口编码 20, NDDL $+0.02$) 估计是视觉可见的。

一旦黑电平设置正确，且可以避免截断信号，对比度则能根据视觉的舒适程度设置为任何级别。在工作室中，对比度的控制可以用来实现参考白色所期望的照度，典型值是 $100\text{cd} \cdot \text{m}^{-2}$ (从历史上看，欧洲使用的参考白色照度稍微低，大约为 $80\text{cd} \cdot \text{m}^{-2}$)。

5.7 对比度和明亮度控制的影响

为了探索对比度和明亮度控制的视觉效果，可以考虑采用一个理想的适应 8 位高清演播室的显示器。

从最佳设置降低明亮度将会导致位于参考黑色以上一点位置的视频内容的裁剪。裁剪不会损害对比度本身，但是在“阴暗处”剔除图像内容将会产生明显的伪影，所以不在降低明亮度上做研究。

计算对比度控制对对比度的影响。假设一个对比度为 3333 (100nt 白色, 0.03nt 黑色) 的典型工作室。减小 20% 的对比度控制降低白色视频信号到 0.8，得到的相对照度为

0.585。增加 20% 对比度控制提高白色视频信号到 1.25，得到 1.71 的相对照度。用 3333 的对比度开始，调整士 20% 的对比度控制，降低对比度到大约 1950 或者将其增加到 5700。

让我们计算对比度控制对“亮度”的影响，用 L^* 估计。调整对比控制士 20% 从而使 L^* 从 81 变化到 118。

计算增加亮度控制对对比度的影响，增加 20% 的亮度控制会使图 5-1 所示的伽马值为 2.4 的幂函数曲线的 y 截距从 -5 变到 +3。参考黑色编码，现在会产生大约为 0.00332 的相对照度；参考白色编码产生的相对照度大约为 1.08。增加亮度控制会使对比度从 3333 降低到 $1/0.00332$ ，也就是到 325。

最后，增加亮度控制 20% 引起参考白色信号 L^* 增加到 103。

增加 20% 的对比度控制会使对比度从 3333 变到 5700，是原来的大约 2 倍。增加 20% 的亮度控制会使对比度从原来的 3333 降低为 325，大约为 9/10 的变化。20% 的亮度控制要比 20% 的对比度控制产生的对比度作用更有效果。

增加 20% 的亮度使 L^* 从 100 变为 103，但是增加 20% 的对比度使 L^* 从 100 变为 118。结果如表 5-1 所示。

表 5-1 调整对比度和亮度的影响

标称值	对比度	参考黑 L^*	参考白 L^*
减少对比度控制 20%	3333	0.3	100
增加对比度控制 20%	1950	0.5	81
增加亮度控制 20%	325	2.8	103

数值例子可由图 5-6 所示的四幅图阐述，展示了对比度（在顶部）和亮度 (L^* ，在底部) 的适应回归控制（在左边）和明度控制（在右侧）的效果，对比度和明度控制范围对应于式 (5-2) 中 m 和 b 的映射。在右上方的图形峰值中通过选择合适的明度控制设置对优化对比度是显而易见的。其他三个图形显示饱和（裁剪），其中在本例中采取设置为工作室视频参考白色水平的 109%，相当于大约为 1.23 的照度。

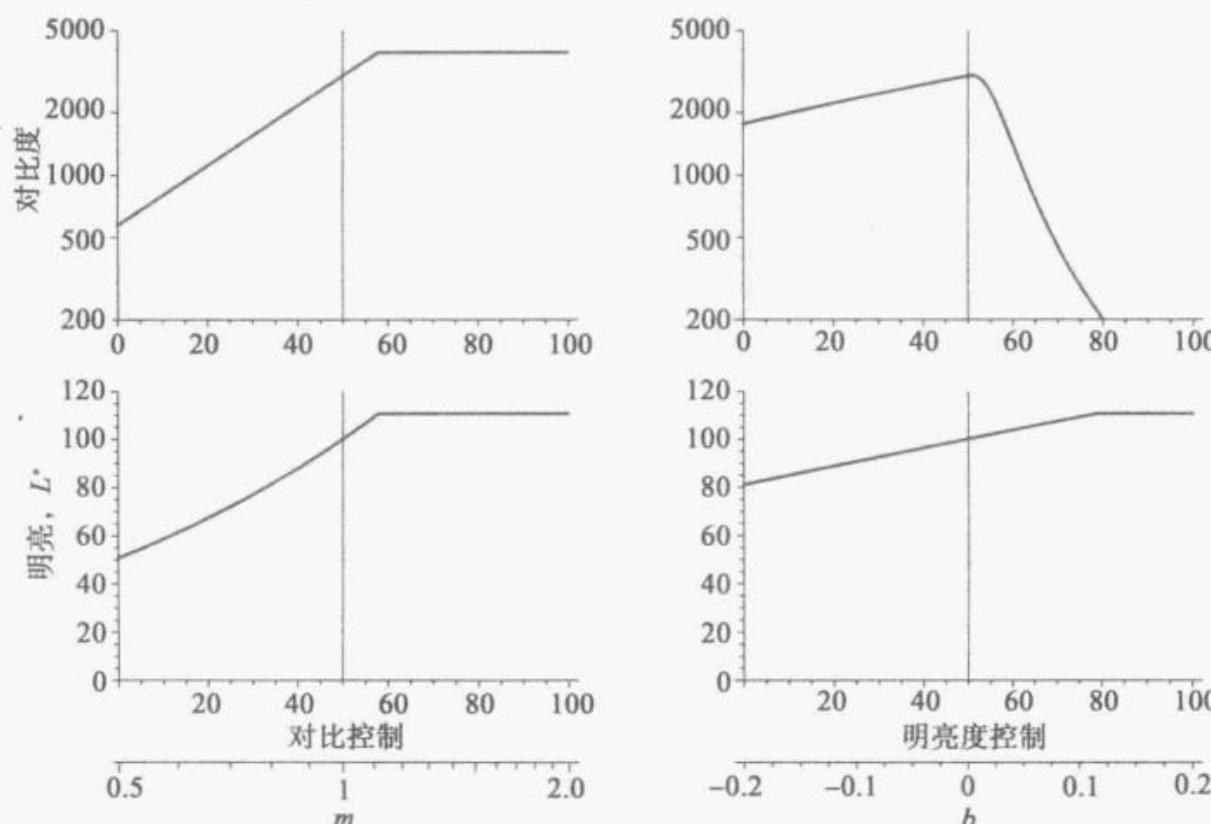


图 5-6 对比度和亮度 (L^*)

两个顶部图的右半侧表明，对高于最佳设置的亮度控制的调整导致对比度的下降速率，大约是调整对比度控制时（在它的非裁剪区）引起对比度变化的3倍：与对比度控制相比，对比度更加适合用亮度控制来调整。从底部图像中可以看出，向上调整对比度控制或者亮度控制来增加白色的明度（直到裁剪），其中的对比控制更加灵敏。

5.8 交替译

在图5-4和图5-5中，当输入一个固定尺度的像素值（归一化的DDL）时，对比度和亮度控制改变的特性。让我们反过来，考虑显示特性是一个固定函数，其显示参考值的变化范围是从0~1。式(5-1)实现图5-1所示x轴的线性操作。对比度和亮度调节可以解释为沿着该轴缩放和偏移。

我们可以建立一个参数B（用户可以访问的黑电平）来控制希望由NDDL0来产生的显示基准值，W（用户可访问的白电平）参数来控制由NDDL1产生的显示基准值。

图5-7所示的背面显示了新的解释。图5-4和图5-5被重新标志显示基准值；下面是像素值（归一化DDL）尺度。DDL的尺度被挤压和偏移。图5-5所示的例子黑电平为0.1，白电平为0.9。黑电平被提升使得NDDL0产生的 L^* 的值大约为3；白电平被设置，以至于NDDL1产生的 L^* 值大约为90。

式(5-1)重新参数化的参数是这样的：

$$y = (W - B) \cdot x + B \quad (5-3)$$

要实现偏移变化适应传统的亮度控制，允许处理具有黑色电平误差的输入信号，B设置应该有一个±0.2的变化范围。为了可以与传统对比控制的增益范围相媲美，W应该从0.5延长到2.0。W的值设置为大于1.1时，大多数显示器将表现出剪裁，有可能需要限制用户设置这样的值。

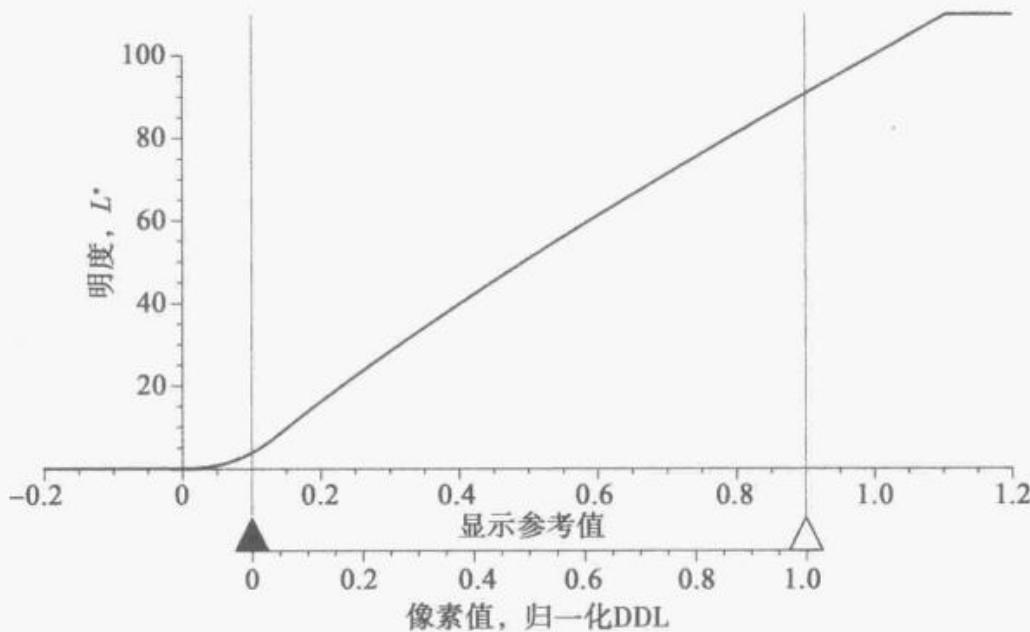


图5-7 黑色电平和白色电平的控制

对于消费类的设备，现代源材料的黑色水平相当稳定，用户的调整补偿差源不再需要。现代显示的周围环境漫反射非常低——大约为0.01——环境照度对对比度的影响非常小。用户调整以补偿环境光不再是必要的。因此制造商应当考虑把黑色电平设置为内部或者服务调整。

在显示端，黑色电平通常用来弥补显示而不是输入信号，所以应该作用在增益（对比度）控制的下游。

在加工设备，有时候有必要在输入信号那里修正黑色电平的错误，同时用保持单位增益：黑色电平控制应该在增益（不应该称之为亮度）应用之前实现。图 5-8 和图 5-9 画出了视频信号路径中对比度和亮度控制的转换方程，忽略了显示器典型的 2.4 幂函数。

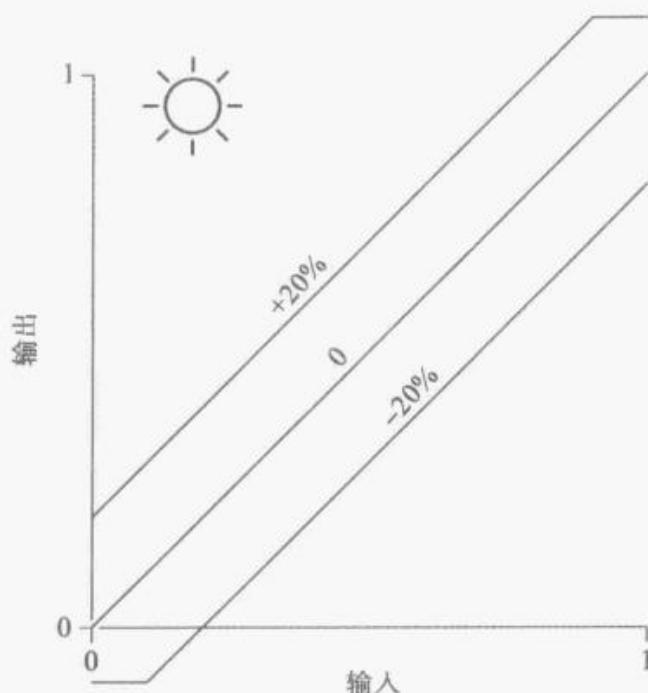


图 5-8 视频中的亮度（黑电平）控制

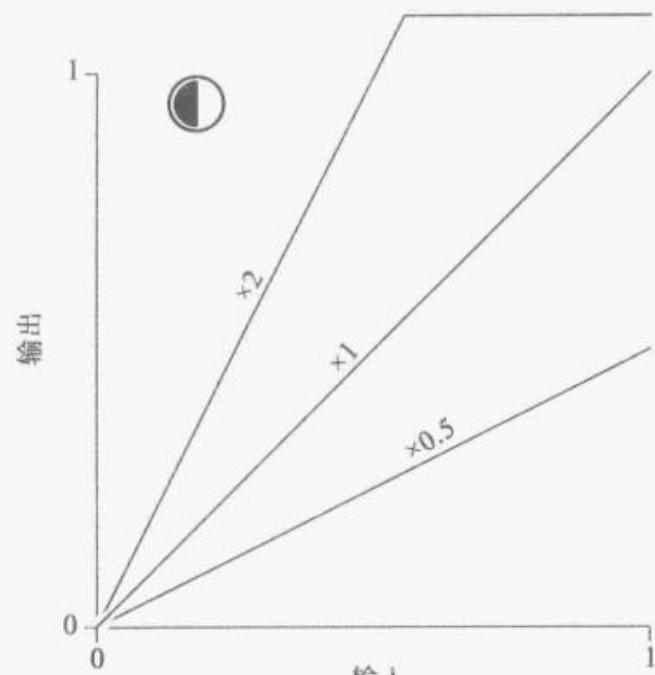


图 5-9 视频对比度（视频电平）的控制

5.9 LCD 中的亮度和对比度控制

在液晶显示器 (LCD) 中，亮度经常会改变背光源的照度；它的作用相当于一个 CRT 显示器的对比度控制。LCD 产生的照度是驱动电压的非线性函数。在早期的 LCD 中，由在面板的偏置电压控制对比度调整。在现代的 LCD 中，信号路径中产生对比度调整增益。对于 LCD，没有什么理由需要分开的 R、G 和 B 的偏置控制 (RGB 低)。

5.10 PDP 中的亮度和对比度控制

在等离子体显示板 (PDP) 显示器中，最大照度是由面板上的电子设计固定的；亮度和对比度控制是由数字信号处理实现的。PDP 显示器产生的照度是驱动电平的线性函数。DDL0 是显示器可能产生的最小照度，所以参考黑色视频编码应该产生 DDL0——没有可能产生其他值。对于 PDP，没有理由产生分开的 R、G、B 偏置控制 (RGB 低)。

5.11 桌面图形的亮度和对比度控制

Adobe 公司的 Photoshop 软件在桌面图形上建立了亮度和对比度控制的实际效果。Photoshop 的亮度控制类似于视频中的亮度控制；然而 Photoshop 的对比度控制与视频中的大有不同。

Photoshop 的控制传递函数曲线如图 5-10 和图 5-11 (相对) 所示。在 Photoshop 中的 R' 、 G' 、 B' 分量呈现给用户的是 0~255 的值。亮度和对比度控制有一个变化范围是土 100 的滑动。

亮度影响 R' 、 G' 、 B' 分量的偏移量在 $-100 \sim +100$ 之间。任何超出 $0 \sim 255$ 的范围的值都会被钳位到最近的极值，0 或者 255。Photoshop 的亮度控制可以与视频的相媲美。但是它的范围（大约是满量程的土 40%）大于典型的视频范围（大约是土 20%）。

Photoshop (传统) 的对比度控制跟随了亮度应用，它适用于增益因子，而不是像

视频对比度控制一样，把参考黑色（零值）固定，Photoshop使得增益围绕着图像数据的加权平均来进行调整，形成 $0.299R' + 0.587G' + 0.114B'$ （具有“伽马修正”视频的图像数据、加权平均值对应BT.601标准的亮度，或者平均像素级——APL）。传递函数，用于对比度调整的各种设置和加权平均值为127.5的图像，这个函数曲线如图5-11所示。

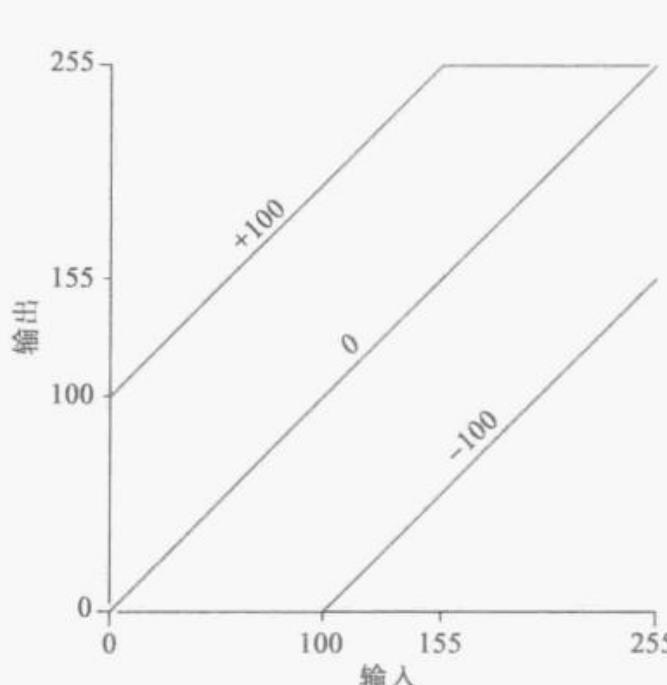


图 5-10 Photoshop 中的亮度控制

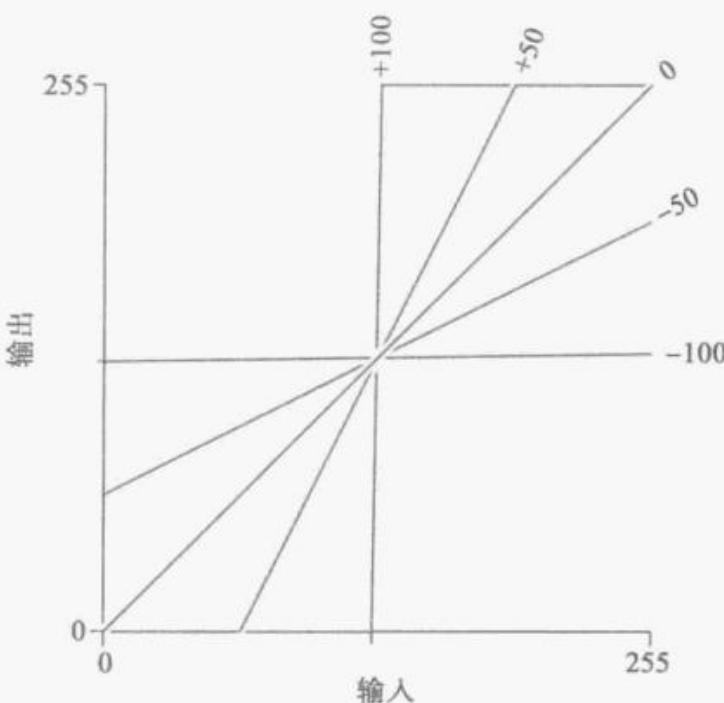


图 5-11 Photoshop 中的对比度控制

Photoshop 的对比度控制的增益变化范围是从零到无穷大，比工作室中典型的 0.5~2 的增益变化范围要大得多。Photoshop 的对比度控制与增益的函数关系如图 5-12 所示。
 $-100 \sim 0$ 的设置增益线性范围从零到单位 1。
 $0 \sim +100$ 的设置，增益是从单位 1 到正无穷的非线性变化，式 (5-4) 描绘了下面的倒数曲线。

$$K = \begin{cases} 1 + \frac{C}{100} & (-100 \leq C < 0) \\ \frac{1}{1 - \frac{C}{100}} & (0 \leq C < 100) \end{cases} \quad (5-4)$$

桌面图形应用程序比如 Photoshop，与视频的 $R'G'B'$ 相比，图像通常是以感知均匀性的方式进行编码的。在一个 PC 上， $R'G'B'$ 分量都是以一个默认的比例再现照度值（三刺激值）的 $1/2.2$ -幂函数。在 Mac OS X 10.6 出现之前的 Macintosh 计算机上，QuickDraw $R'G'B'$ 分量与显示照度（或者三刺激值）的 0.66 幂函数成默认比例。现代 Macintosh 计算机符合 sRGB 标准。然而，在 PC 和 Macintosh 计算机中，用户、系统软件或者应用软件可以设置传递函数为一个非标准函数——或许就影响了线性光编码——正如将要第 27 章中描绘的那样。

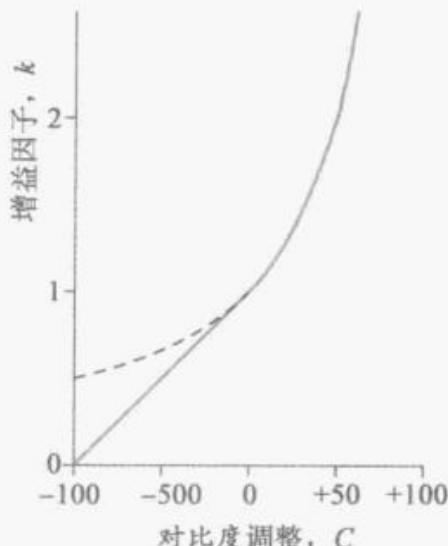


图 5-12 Photoshop 中的对比度控制增益因子

第 6 章

计算中的光栅图像

本章节将视频放置在计算的背景中。计算中的图像分为三种形式，图 6-1 所示的为计算机图像被原理性地描绘为三层：特征图像描述，光栅图像，以及压缩图像。

特征图像描述的不是直接包含一个图像，而是包含一个二维或者三维的高层次的几何描述，例如，它们的对象和属性。这种形式的二维图像有的时候叫做向量图像，虽然它的原始图像比直线段的向量表示要复杂得多。

光栅图像以扫描线的顺序直接列举了每个像素的灰度或者色彩内容。有四种光栅图像的基本形式：二值图像，伪彩色图像，灰度图像和真彩色图像。图 6-1 所示的四种形式的图像类型从左边的低质量到右边的高质量按列排列。

压缩图像起源于光栅图像数据，但是数据已经被处理，以减小存储或者传输要求。图 6-1 底部的一行包括若干个压缩方法，左边是无损（数据）压缩方法，一般适用于二值图像或者伪彩色图像数据，右边是有损（图像）数据的压缩方法适用于灰度图像或者真彩色图像。

对于灰度图像、伪彩色图像和真彩色图像，在计算系统中引入查找表（LUT）的方法来描述图像，这个表是把像素值映射到显示 $R'G'B'$ 值。大部分计算系统采用均匀感知的图像编码，有一些系统会采用线性光编码，或者采用其他的技术。对于采用均匀感知方法的系统，类似于或者兼容在视频中，它的 LUT 的方式需要导入适当的传递函数。如果 LUT 载入的传递函数使得编码值与强度成比例，那么感知均匀性的优势将会减弱或者消失。

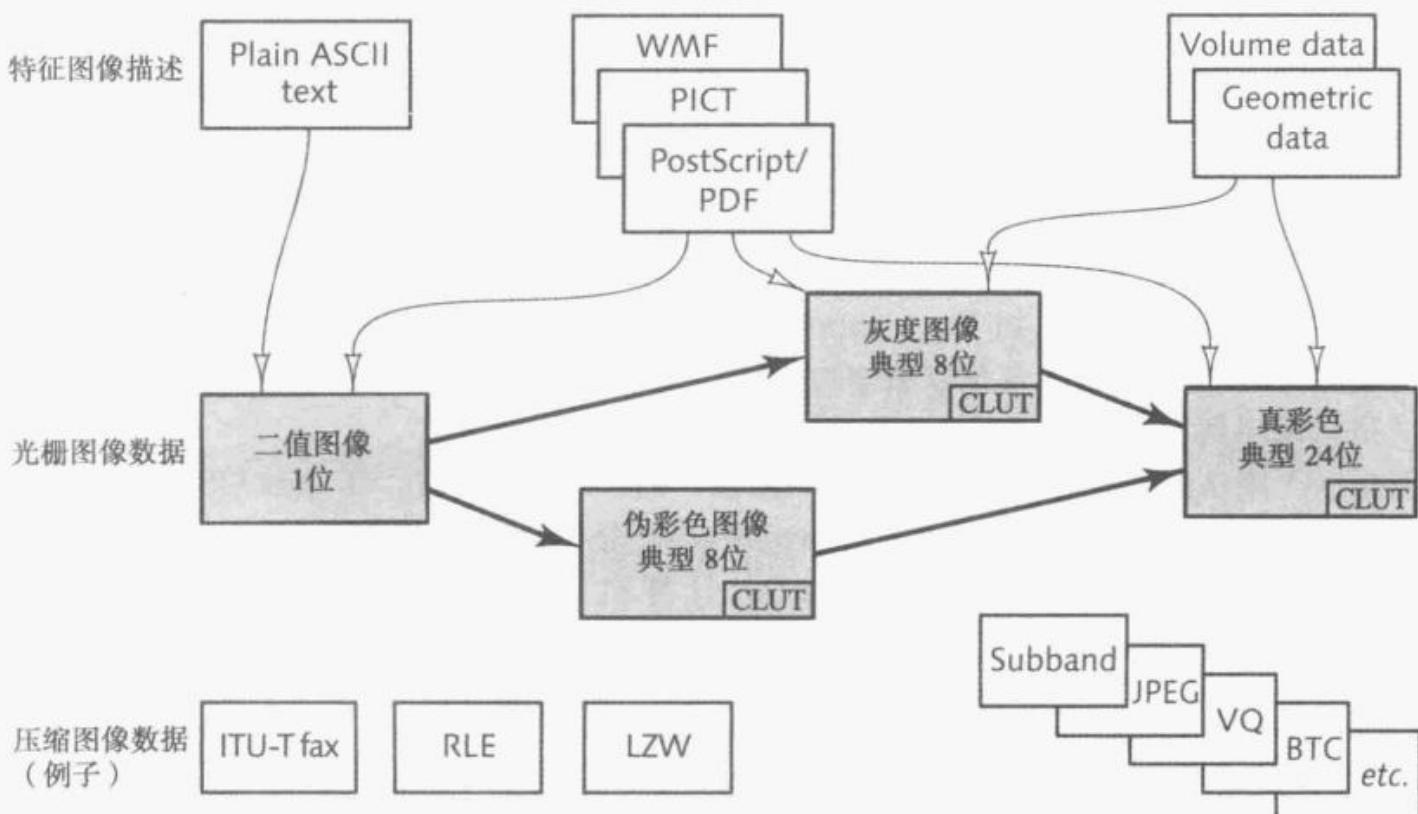


图 6-1 光栅图像数据

许多不同文件在使用时都使用了这种陈述，讨论文件的格式超出了本书的范围。传输照相质量的彩色图像，文件的格式必须满足每个像素至少 24 位。为了采用最大限度感知地利用每个分量的有限位数，应该采用非线性编码，就如本书 3.3 节所述。

6.1 特征图像的描述

许多方法被用来描述一幅图像的内容，这种方法的抽象级别比直接列举的级别更高。象征图像数据采用光栅化过程转换为光栅图像。图像的光栅化（成像或者渲染）是通过解释符号数据和生成光栅数据来实现的。如图 6-1 所示，这种操作过程从顶行到中间行传递信息。

几何数据描述对象的位置、大小、方向和其他属性；三维几何数据可以解释为从一个特殊的观察点产生图像。对几何数据进行栅格处理叫做渲染；真彩色图像经常这样产生。

Adobe 公司的 PostScript 系统广泛用来表示二维的插图、排版元素和出版物。PostScript 本质上是一种编程语言，专门用于成像操作。当一个 PostScript 文件由 PostScript 解释器执行时，这个图像就被呈现了（在 PostScript 中，光栅化操作经常叫做光栅图像处理或者翻录）。一个压缩的 PostScript 文件（EPS 或者 EPSF）描述了一幅图像的 PostScript 文件的特殊情况；Adobe 的 PDF 格式实际上是 PostScript 中一个不可编程的变体。

一旦栅格化，一般就会没有办法把光栅图像数据转换到特征性描述的数据：光栅图像——图 6-1 所示的中间行——一般不能返回到它最顶行的描述。如果您的应用程序涉及渲染图像，你会发现保留特征数据仍很有用，即使是在渲染之后，若需要以不同的大小或者执行诸如删除一个对象的修正重新渲染图像，仍然可以做到。

从一台传真机，一个摄像机，一个灰度或者彩色扫描仪出来的图像起源于光栅图像格式：没有特征描述可以用。用光学字符辨识（OCR）和光栅到向量技术尝试从光栅图像中提取文本或者几何数据，但是结果一般不令人满意。

6.2 光栅图像

有四种类型的光栅图像数据如下。

- 二值图像，顾名思义每个像素有 1 位；
- 灰度图像，通常每一个像素 8 位；
- 伪彩色图像，通常每个像素 8 位；
- 真彩色图像，每个像素一般是 24 位。

灰度和真彩色系统能够代表连续色调的系统。视频系统只采用真彩色图像（灰度图像是一个特例）。

在下面的章节中，将解释二值、灰度、真彩色，以及伪彩色。真彩色和伪彩色的描述都是伴随着代表帧缓存和显卡后端硬件的框图来进行的（历史上，这个硬件将包括数/模转换器（DAC）；现在，数字显示器接口使用例如 DVI、HDMI，以及 DisplayPort）。或者，也可以考虑每个框图来代表一个算法，将图像数据转换为显示 $R'G'B'$ 各个分量。

6.2.1 二值图像

每个二值图像（或者两电平）的像素包括 1 个位，分别代表黑色或白色/（但是没有二者之间的值），在计算中这通常叫做单色（该术语应该表示一个单一的色调的阴影，然而在普通的使用中——尤其在视频中——单色表示一个图像的黑与白，或者灰度，图像的分量）。

由于数据通信的发明，二进制数 0 一直称为空间，二进制数 1 一直称为标记。电子显示装置上的“标记”发射光，所以在视频和计算机图像学中二进制数 1（或者最大编码值）通常表示白色。在印刷中，一个“标记”把墨水留在页面上，所以在打印中二进制数 1（或者在灰度中，最大像素值）通常代表黑色。

6.2.2 灰度图像

灰度图像代表色调有效的连续范围，从黑色，经过中间的灰色阴影到白色。每个分量有足够的数据位的灰度系统，8 位或者更多可以表示一个黑色和白色的照片。一个灰度系统或许有或许没有 LUT；它可能也可能不会均匀地感知。在印刷技术中，一个灰度图像是有连续色调的（从艺术线条或者类型区分）。当一个连续色调的图像被打印时，通常会使用半色调。

6.2.3 真彩色图像

真彩色系统的每个像素具有单独的红色，绿色和蓝色分量。大多数真彩色系统每一个分量由 8 位的字节表示（称为每个通道 8 位，或者位每通道（*bpc*））。每个像素有三个分量（通道），因此这种模式称为“24 位色”（或者“百万种颜色”）。每个像素的 RGB 值的不同编码可以代表 2^{24} 或者 16.7×10^7 。在计算中，一个真彩色缓冲区通常有三个 LUT，每个分量一个。24 位真彩色系统的 LUT 和数/模转换器（DAC）如图 6-2 所示。

从图像编码值到电压显示的映射是由 LUT 的内容来确定的。由于图像的感知均匀性，当每一个 LUT 包含的标识函数（“斜坡”）映射着的输入与输出不变时，最佳使用感知一般是由真彩色像素值组成的。

在计算中，LUT 可实现从编码值到三刺激值（同样，对于强度）的任意映射。表示可以区分的颜色的像素值总数取决于所使用的传递函数。如果 LUT 实现了一个幂函数来施加在线性光数据中的伽马修正，编码 100 的问题会出现最坏的情况。具有 24 位色彩和正确选择的传递函数，照片质量的图像可以被显示，几何目标可以顺利地渲染，对于很多应用具有足够高的质量。但是如果 LUT 用来设置每个分量用 8 位表示的线性光，则在许多图像中会明显地显示轮廓，正如本书 3.3 节提到的那样。拥有 24 位真彩色图像不是良好图像质量的保证。如果一台扫描器声称每个像素有 30 位（或者 36 位），很明显每个分量有 10 位（或者 12 位）。然而，这些值是否是物理（线性光强度，不严谨地称为“强度”）编码或者感知（作为一个与明度可比的数量）编码具有很大的区别。

在视频中，不管 LUT 是否存在，或者各自设置为一致函数。在有感知统一编码的真彩色模式中演播室的视频系统得到有效的永久连接：编码值直接呈现给数/模转换器，不用插入 LUT。

当每个像素具有的字节数是 2 的幂时，设计一个帧缓冲存储器系统是最简单的；因此，一个真彩色帧缓冲区每个像素通常有四个字节——“32 位色”，三个字节组成红、绿、蓝分量；第四个字节用于其他目的，而不是代表颜色。第四个字节可以包含叠加信

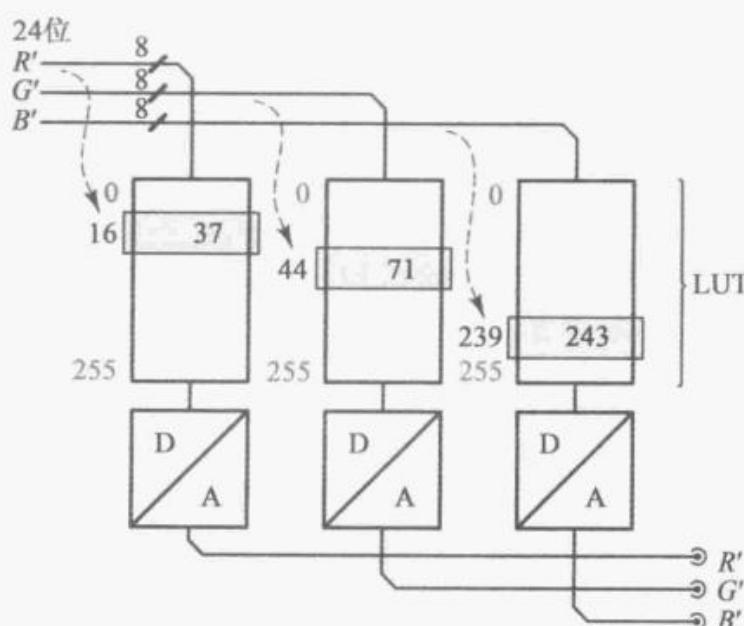


图 6-2 真彩色图像（24 位）系统的 LUT

息，可以保存 α 或者关键分量表示从黑色（0，完全透明）到白色（1，完全不透明）不透明度。在计算机图形学中， α 分量调制元件（通常是 RGB）在线性光领域编码。在视频中，线性关键信号调制非线性（伽马修正） $R'G'B'$ 信号或者亮度和色差分量，例如 $Y'C_B C_R$ 。

6.2.4 伪彩色图像

在伪彩色（或者索引颜色，或者颜色映射）系统中，伪彩色用位来表示——通常是 8 位，组成一个图像或者帧缓冲的每个像素。这为每个像素提供了独特的编码值——通常为 256。在颜色数量相当少的地方，伪彩色涉及“通过数字绘画”。在一个 8 位的伪彩色系统里，任何特殊的图像或者帧缓存在任何时刻的内容，都是限定在所有可用颜色中的 2^8 （或者 256）种颜色中。

每个编码值在 LUT (CLUT, 颜色映射或者调色板) 中用作一个索引来检索 $R'G'B'$ 分量；数/模转换器 (DAC) 线性地转换这些电压值，并应用到显示器中 (Macintosh 是一个例外：从 CLUT 读出的图像数据通过第二个 LUT 起作用)。伪彩色的 CLUT 值被有效地感知编码。

一个 8 位的伪彩色系统的 CLUT 和数/模转换器 (DAC) 如图 6-3 所示。一个典型的 LUT 检索每一个红绿蓝 8 位的值，所以每 256 个不同颜色可以从总共的 2^{24} ，或者 16 777 216 种颜色中选择。(一个 CLUT 对每个分量可以返回 4、6 位或者多于 8 位。)

伪彩色图像的数据总是伴随着相关的颜色映射（或者调色面板）。颜色映射可以是固定的，独立于图像，或者是针对特定的图像（自适应或者优化）。

固定 CLUT 的流行选择是浏览器安全调色板包含 216 种颜色，这些颜色选自集合 $\{0, 51, 102, 153, 204, 255\}$ 的 8 位 R', G', B' 组成。这一套 216 种颜色非常适合于 8 位伪彩色 CLUT；色彩被感知地分布在整个 $R'G'B'$ 立方体里。

伪彩色图像适用于如地图，原理图或者卡通片类的应用，对于这种图，在图像的任一点，每个颜色或者其组合可以是完全存在或者不存在。在一个典型的 CLUT 中，相邻的伪彩色编码一般是完全无关的；例如，分配给编码 42 的颜色与分配给编码 43 的颜色没有必然的联系。

6.3 不同类型之间的转换

在图 6-1 中，左边到右边对应于无损失的转换。

现在先不管伪彩色图像，在图 6-1 中其他三个方案的数据，只要适当地分配正确的码值，可以扩大到任何方案。例如，一个灰度图像可以通过从黑色到白色分配编码来扩展为彩色。扩大增加了位数，但没有增加信息。

一个伪彩色图像可以通过 CLUT 的软件应用转化为真彩色图像。转换为真彩色图像可以实现无损失，但是前提是真彩色的 LUT 是合理的。

关于相反方向的转换，当且仅当它包含的颜色或者形状是图 6-1 所示模式的左边，图像才可以无损失地缩小。否则转换时会引入图案或者颜色的损失。

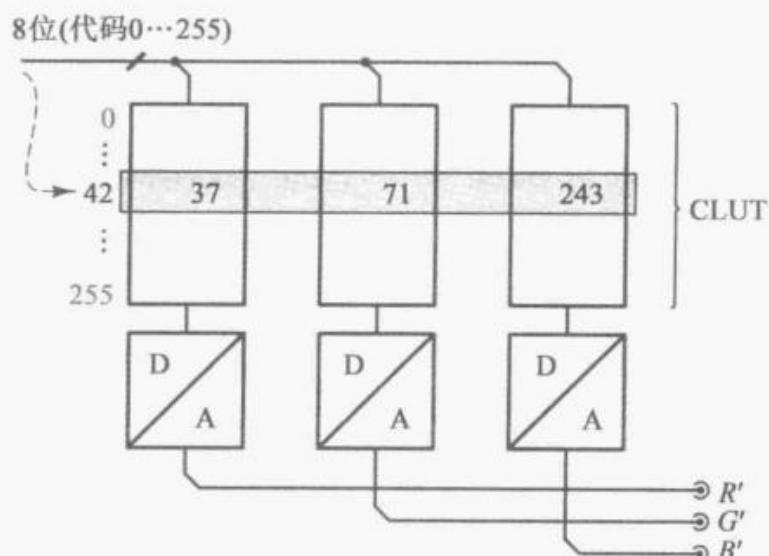


图 6-3 伪彩色 (8 位) 图像的 CLVT 和 DAC

一个伪彩色图像可以通过一个固定颜色映射的软件应用逼近真彩色图像。同样，一个颜色映射的量化算法可以用于检查特定的图像（或者图像序列），并且计算可以优化或者适用该图像或者序列的颜色映射。

6.4 图像文件

二值、灰度、伪彩色、真彩色格式的图片可以存储在文件中。通用的图像文件格式在其头信息中存储像素的行数和列数。

一些文件格式——例如 TIFF 和 EPS——存储了有关图像的大小信息。预期图像的宽和高可以直接以绝对单位如英寸或者毫米的形式来存储。另外，该文件可以以每英寸的像素数 (ppi)，或者不太准确的，每英寸的点数 (dpi) 为单位来存储样品密度。样品密度通常会混淆地称为“分辨率”。

在一些软件封装中，例如 Adobe Illustrator，文件中的预期图像的尺寸编码受到关注。在其他的软件中，例如 Adobe Photoshop，不管在文件或者显示器中每英寸像素的数目，100% 比例观看意味着在文件像素和显示器像素之间有 1 : 1 的关系。没有尺寸信息的图像文件通常视为每英寸有 72 个像素；不知道影像尺寸信息的应用软件经常使用 72ppi 的默认值。

6.5 计算机图形学中的“分辨率”

在计算机图形学中，一个像素通常视为强度分布均匀地覆盖在屏幕中一小块区域。在固定像素的显示器，例如液晶显示器 (LCD)，等离子显示面板 (PDP)，数字光处理 (DLP) 显示器，这些离散像素分布在显示器装置上。当这些显示器以原像素数被数字驱动时，在帧缓冲区像素和设备像素之间有一对一的关系。然而，当面对帧缓冲区像素计数和显示设备像素之间不匹配时，图形子系统可能用原始的手段进行重新采样。如果帧缓冲数较高，像素就会被丢弃，如果较低，像素值就会被复制。在这两种情况下，图像质量就会受到影响。

CRT 显示器每个像素的光通常具有高斯分布，这将会在下一章中讨论。典型的光点尺寸是这样的，在相邻的像素分布有一些重叠。你可能会认为通过相邻的显示单元产生的光分布的重叠是不需要的。然而，图像显示要求重叠要有一个确定的度，目的是最小化像素结构和扫描线结构的可见度。本书将在图像结构中讨论这个问题。两个不同的措施称为计算分辨率。

图像的列和行数——也就是像素的行数和列数——在一个帧缓冲中。

用于图像数据（每英寸代表的点数，dpi）的每英寸的像素数 (ppi)。

图像科学家认为应当交付给观众的解决方案是：从显示器所显示信息表面（或者屏幕本身）适当地进行分辨率估计。这两项措施最重要的是限制了分辨率，但是没有一个直接量化了分辨率。在第 9 章，将会介绍这个词如何在图像科学和视频领域使用。

第 7 章

图像结构

数字图像处理简单的方法是把一幅图像当做独立像素的矩阵，而忽略每个像素的光功率在空间的分布。你可能会认为当相邻两个像素之间没有重叠时便能获得最佳图像质量；许多计算机工程师持有这个观点。然而，连续色调图像在像素之间最好具有一定的像素重叠；清晰度稍微有所降低，但是像素结构却不可见，图像质量也提高了。

显示像素之间的强度分布称为点扩散函数（PSF）。通过 PSF 中心的一维切片俗称一个点的轮廓。显示器的 PSF 影响再现图像的本质。PSF 的作用可以使用过滤理论来解释，第 20 章将讨论一维的情况，第 22 章讨论二维的情况。

从历史上看，灰度（“黑与白”）CRT 显示器的 PSF 大致是高斯分布形状：强度峰值分布在像素的中心，在小范围内下降，在一定程度上重叠了相邻的像素。彩色的 CRT 的扫描点也有这样的形状，但是 PSF 受荫罩和荫栅的影响。引入直接观测彩色 CRT，把空间滤波的需求转移给了观众：假设观众距离屏幕足够远，以至于观众的视线系统会进行空间整合，进而对三颜色结构进行模糊处理。

现代直视固定像素显示器（FPD）例如 LCD 和 PDP 显示器对应于每个颜色分量（子像素）的大部分面积都有较为一致的光发射；这个调制光的空间结构与直视彩色 CRT 相媲美，同样取决于站在足够远的观察者的视觉系统，可以进行空间一体化来模糊三颜色结构。

一个分布在屏幕上一个小的方形区域的光强均匀的像素有一个称为“箱（box）”的点扩散函数。

7.1 图像重建

图 7-1 再现了一个理想化位图（两个电平）图像的一部分，这是计算机显示桌面的一部分。每个样本或者是黑色或者是白色。具有水平“条纹”的元素是窗口标题栏的一部分；棋盘背景的目的是整合到黑色。图 7-1 展示了具有“盒子”分布的图像的重建。每个像素都在其范围内均匀着色；各个像素之间没有重叠。这个图充分体现了显示在 LCD 上光栅锁定的图像。对于光栅锁定，我指的图像数据是与像素阵列对准的底层图像元素。

CRT 电子枪产生一个照射到荧光屏斑点的电子束。光束偏转，从而形成一个用来追踪整个屏幕扫描线的栅格图案，本书将会在以下章节展开描述。当光束从电子枪发射时不能准确聚焦，在传输至荧光屏的过程中进一步发散。在屏幕的表面每个像素所产生的强度具有类似二维高斯函数的“钟形”分布。在一个光斑重叠的典型情况下，本实例的方格区将会显示为几乎均匀的灰度，如图 7-2 所示。你可能会



图 7-1 位图图像的箱重建

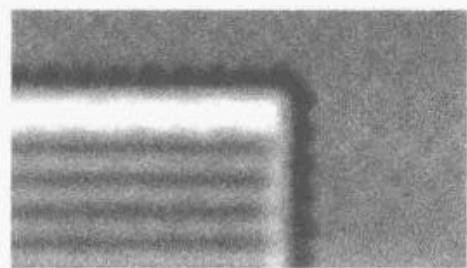


图 7-2 高斯重建

认为重叠造成像素的模糊会降低图像的质量。然而，对于连续色调的图像（“连续色调”），一定程度的重叠不仅仅是可取的，有的时候是必要的，你将会在下面的例子中看到。

图 7-3 显示了一个 16×20 像素值的图像，它的暗线比一个像素略宽，且 7.2° 垂直倾斜。左边的图像数据采用箱分布进行重构，锯齿状和质量差的特点是显而易见的。在右边，图像数据使用高斯分布进行重建。结果是模糊的，但锯齿程度轻。

图 7-4 展示了用两种方法来重建相同的 16×20 像素 (320B) 的连续色调的灰度图像数据。左边的图像使用一个“箱”函数进行重建，右边的图像使用高斯分布方法。这个例子重建后，每个图像是 4cm (1.6in) 宽度的构造。典型 40cm (16in) 的观看距离，像素对向角是 0.4° ，此时视力灵敏度接近其最大值。在这个距离，当采用箱函数进行重建时，每个图像的像素结构是非常明显的；像素结构的可视性超过了图像本身的感知。合适的图像是采用高斯分布重建的。虽然模糊，可是很容易识别作美国文化标志。这个例子说明清晰并不总是好的，模糊并不总是坏的！



图 7-3 对角线重建



图 7-4 连续色调图像重建



图 7-5 显示了一个包括 20 份图 7-3 (左边) 所示，顶行的 16×20 像素的图像。考虑一个 20 帧动画序列，每一个帧是由图 7-3 中的一系列图像的行形成的。这个动画将描绘一个狭窄的垂直线，以每 8 帧一个像素的速度向右穿过屏幕。如果使用了图 7-3 (左边) 所示的图像行，移动线的宽度会出现帧到帧的抖动，最小的明度会发生变化。通过高斯分布重建，如图 7-3 (右边) 所示，运动描述更为光滑。

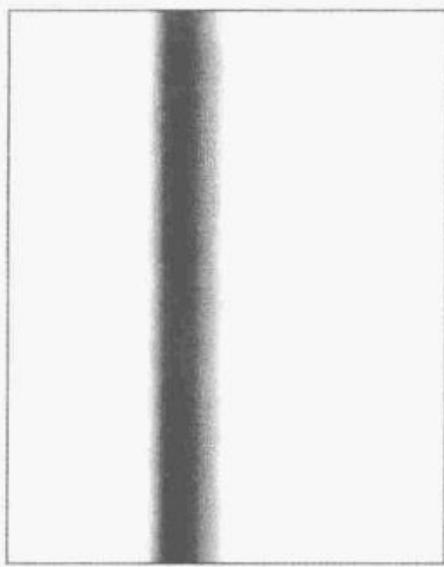


图 7-5 动画序列的一帧

7.2 采样孔径

在一个实际的图像传感器中，每个元素从图像平面的有效区域获取信息；每个像素值是强度在该区域的分布函数。一个图像捕获装置的像素灵敏度的分布称为采样孔径，一种逆排序的 PSF——你可以把它称为一个点“集合”的函数。采样孔径影响了源于传感器的图像信号的性质。在连续色调成像系统中，采样孔径通常在像素的中心出现峰值，并在较小的距离内下降，在一定的程度上重叠相邻像素。

1915 年，哈里·奈奎斯特发表了划时代的论文，指出一个采样信号不能被准确地重建，除非所有的信号的频率都在采样频率的一半以内。这个条件后来称为奈奎斯特采样定律；采样速率的一半称为奈奎斯特速率。奈奎斯特在一维信号中发展了他的理论，但是现

在已经扩展到了二维。在一个数字系统中，至少有两个元素——两个像素或者两个扫描行——来表示一个周期。一个周期相当于一个片子的线对，或者两个“电视线”(TVL)。

在图 7-6 所示模式中，黑色的正方形被规则排列的孔打穿，这些孔代表小型采样孔径的网格。采样网格的背后是一组十二个的黑条，偏离垂直方向 14° ，代表图像的信息。在图像的采样区域，你将会看到倾斜 45° 的三条宽黑条。这些条代表空间混叠现象，因为图像中每英寸的条数大于每英寸采样格子数量的一半数量。混叠是可以避免的——或者可以最小化——通过在采样之前强加一个空间滤波器即可实现，本书将在第 20 章描述一维信号，在 22.4 节描述二维情况。

奈奎斯特解释任意一个信号可以准确地重建，只要最高频率分量的采样信号多于两个。应用于图像，单位距离的样本数至少 2 倍于图像元素的数量。图 7-1 所示的棋盘形图案无论在垂直还是水平维度上都不符合这个标准。此外，标题栏在垂直上不符合标准。这样的元素只有在精确注册——“锁定”——到成像系统的采样网格时才可以在一个二值图像上表示。然而，从实际中拍摄的图像几乎从来没有恰好与网格对齐的元素！

点采样是指使用无穷小的采样孔进行捕捉的方法。在连续色调的图像中这是不希望出现的。如果一个像如图 7-1 所示的物理场景被旋转 14° ，使用点采样机进行拍摄，“箱”分布进行显示，图 7-7 表明了将会出现什么情况。交替开关元素在棋盘部分和标题栏上呈现混叠（混叠现象是显而易见的，即使图像是使用高斯分布重建的）。这个例子强调，在数码图像中，我们必须代表任意场景，而不仅仅是元素与采样网格具有亲密关系的场景。

合适的预采样滤波器可以防止（或者至少减少）图 7-6 所示的莫尔条纹伪影，防止或者最小化图 7-7 所示的混叠。当包含图 7-2 所示的标题栏例子和桌面图案的图像内容呈现给预采样滤波器时，会产生模糊。只考虑例如图 7-1 所示的位图图像，你就可能会觉得模糊是有害的，为了避开在拍摄高品质连续色调的图样时空间走样，在相邻传感器元件的灵敏度分布的一些重叠是必要的。

上面介绍了来自于较差的 PSF 捕捉的混叠伪影，我们现在可以回到显示并且讨论重建 PSF（光斑轮廓）。

7.3 光斑轮廓

连续色调图像显示系统设计者的目的是使显示屏允许在宽图像视角观看，同时具有最小伪影的干扰，例如，混叠现象或者可见的扫描线或者像素结构。图片大小、观看距离、光斑轮廓和扫描线或者可视像素都是相互作用的。显示系统设计者无法直接控制观看距离；光斑轮廓是可以优化的参数。

在本书将会演示“箱”分布和高斯分布的不同。图 7-3 和图 7-4 显示周边分布有重叠是可取的，甚至当重构图像在近距观察时，有明显模糊的情况也是可取的。

当图 7-3 或者 7-4 所示的图像在 10m (33 英尺[⊖]) 的距离观看时，像素对向角度是 1 弧分 ($(1/60)^\circ$)。

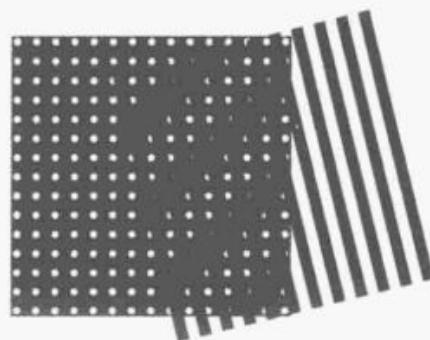


图 7-6 莫尔模式



图 7-7 旋转的位图图形图像

[⊖] 1 英尺 (1ft) = 30.48cm。——编辑注

在这个距离，由于人类的视觉灵敏度有限，两个图像显然是相同的。想象一下，在这些图像旁边放置具有无穷小点的发光显示器，产生的总光通量与完美的白色像素相同。在距离 10m 时，发光显示器的像素结构会有所可见。在一个较远的观看距离——比如像素或者扫描边线的对向角度小于 $(1/180)^\circ$ ，对应于在 3 倍的正常距离或者大约 $20 \cdot PH$ 观察的标清——人的视觉系统的灵敏度有限，这导致所有的三个显示器看起来相同。随着观察者的移近，取决于光斑轮廓的不同的效果开始变得明显。我们将讨论这两种情况：“箱”分布和高斯分布。

7.4 “箱” 分布

一个典型的数字投影机——例如 LCD 或者 PDP——具有光斑轮廓，其类似于一个覆盖在像素间距的几乎整个宽度和高度的“箱”分布。在图像的行与列之间没有显著间隙。每个像素有三个颜色分量，不过，投影装置的光学系统设置成由分量的光分布叠加的。在较远的距离，像素结构不可见。然而，随着观看距离的减小，将会引入混叠（“锯齿”）。投影镜头的有限性能可以在一定程度上缓解混叠；然而，混叠可以是相当明显的，如图 7-3 和图 7-4 所示例子。

一个典型的直接观看的数字显示器，如 LCD 或者 PDP，每个像素包括三个颜色分量，这些分量对应于每个像素的不同区域。通常这些分量是并排的。在图像的行之间没有显著的间隙。然而如果一个分量（如绿色）接通，其他的关掉，则列之间就会有间隙。这些系统依赖于观察者有限的灵敏度去组合这些分量到单一的颜色区域。在一个较近的观察距离，间隙是可以见到的，这可以引起混淆。

使用“箱”分布方式显示的观看距离，如直接观看 LCD 或者 PDP，受到引入混叠的限制。

7.5 高斯分布

正如刚刚所讲，CRT 显示器有一个类似于高斯分布的光斑轮廓。

对于高斯分布小型光斑，光斑宽度小于 $1/2$ 的扫描间距，线结构即使在一个非常大的观看距离上也会变得明显。

对于高斯分布的中型光斑，斑点的宽度约等于扫描线的间距，相比小型光斑，在较近的距离扫描线会变得可见。

光斑的尺寸增加到超过两倍的扫描线间距时，由于光变得这样大，以至于在线结构之间没有进一步的改善，可视性变得更大。然而，让光斑变得大于所需时，有一个严重的缺陷就是：锐度降低。

在本章的开头，在图像显示中为了避免显示像素结构，在相邻显示元素产生光的分布有一定的混叠是必要的。这样的混叠降低了锐度，但是降低了多少？需要多少混叠？本书将会在第 9 章进行讨论。尽管如此，首先，本书首先会介绍光栅扫描基础。

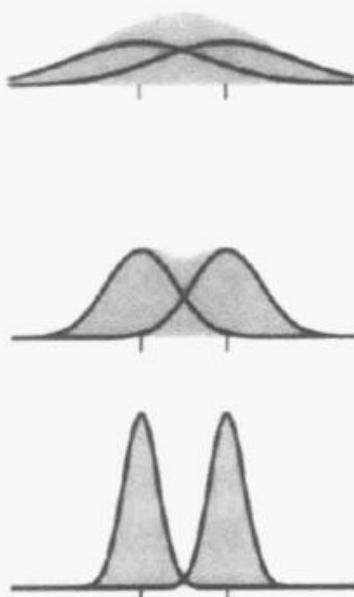


图 7-8 高斯光斑尺寸

光栅扫描

前文已经对像素阵列做过介绍。本章概述光栅扫描的基础过程，这里像素阵列的样本均匀地按时间排列，以形成扫描线，这些扫描线又及时贯穿于各帧的时间间隔。在标清分量介绍中，我将详细地介绍在传统的“525 行”和“625 行”的视频扫描。对于复合 NTSC 和 PAL 制式，我将介绍在这些系统中使用的颜色编码。第 15 章将介绍高清晰度电视的扫描方法。

8.1 闪烁率、刷新率和帧频

一系列的静止图像，如果被捕捉并以足够高的速率显示，可以产生运动的错觉。

传统的用于电视的 CRT 显示器，发光时间只占据帧时间的一小部分：该显示器具有短的工作周期，大部分时间它是黑的。如果闪烁率或刷新率太低，闪烁会被察觉。视觉的闪烁敏感度取决于显示器和观察条件：为了避免闪烁，必须有明亮的环境，较大的图片观察角度和较高的闪烁率。因为图像观察角度影响闪烁，闪烁取决于观看距离。

最现代化的显示器包括液晶显示器和等离子显示板，有不闪现、不闪烁的特点。尽管如此，它们可能会表现出各种运动画面缺陷。

在闪烁的显示器上，所显示的图像本身的明亮度在一定程度上会影响闪烁阈值，所以图像越亮就需要更高的刷新率。在非常暗的环境中，诸如电影院，闪烁灵敏度完全是由图像本身的照度来确定的。余光比中央（中央凹）区域拥有更高的敏感性，所以闪烁门槛的提高在一定程度上依赖于更宽的视角。表 8-1 总结了在电影、视频和计算机使用的刷新率。

表 8-1 刷新率

应 用	显 示 照 度	周 围 环 境	刷 新 (闪 烁) 率 /Hz	帧 频 /Hz
电 视	48nt	黑~0%	48	24
	80nt	暗淡~5%	50	25
	120nt	暗淡~5%	≈60	≈30
办 公 室	320nt	“平均”~20%	各种速率，如 66、72、76	与刷新率一样

表 8-1 刷新率是指整个画面更新的最短间隔。闪烁率是指画面覆盖在显示器上的速率。不同的刷新率和闪烁率用于不同的情况。

处于电影院的黑暗环境中，48Hz 的闪现速率足以克服闪烁。在电影出现的初期，每秒 24 帧被认为是足够好的运动呈现效果。因此，传统的胶片投影仪采用了双叶片快门，如图 8-1 所示，每帧闪烁两次。更好的效果能够通过每秒 48 帧的设备实现，或可以用单刃快门来获得，但这样的做法不符合标准。

在观看电视的昏暗环境下，如客厅，60Hz 闪烁率就足够了。8.4 节讲述的隔行技术使得视频拥有一个与电影放映机的双叶片快门相媲



图 8-1 双叶片快门

美的功能：每个帧分为两个字段。刷新取决于场速率（2 倍的帧频）。对于一个给定的数据速率，隔行技术使闪烁率加倍，并通过加倍时间采样率提供改进的运动影像。不含隔行技术的扫描称为渐进。

在办公环境中使用的 CRT 显示器所需的刷新率通常都高于 70Hz 来克服闪烁（见法雷尔，1987）。具有不闪烁特点的 LCD 显示器现在已经取代了 CRT 显示器。不再需要高刷新率来避免闪烁。

8.2 扫描简介

之前我提出电视刷新率是如何选择才能尽量减少闪烁。在 9.3 节中，我将概述空间采样如何确定图像阵列中像素的数量。视频扫描表示有序的像素，以便获取、输送、处理，或显示与每个帧相关联的固定时间间隔的每一个像素。在模拟视频中，在一个固定的短的时间内——有效行时间，图像平面上的信息以均匀速率从左向右扫描。扫描建立了图像中的位置和信号的时刻之间的固定关系。连续的行是以均匀的速率从顶部到图像的底部进行扫描的，所以在垂直的位置和时间之间也存在固定的关系，设置于图像中的平行扫描线的固定模式是光栅。

数字图像阵列样本通常是以与常用的输送模拟视频相同的顺序传送的：首先传输最上面的图片行（从左到右），然后连续行。扫描线是一个老式的术语，现在图像行是比较常用的说法。连续像素处于图像列中。

在摄像机和显示器中，一定的时间间隔消耗在改善行与行之间的扫描操作上，历史上称为横向回扫；垂直回扫需要消耗几个行扫描时间（从一个扫描的结束到一个扫描的开始）。在这些间隔中 CRT 的电子枪必须处于关闭（清空）状态，所以这些区间称为消隐间隔。横向消隐出现在扫描线之间；垂直消隐间隔（VBI）发生在帧（或场）之间。图 8-2 显示了“525 行”的视频消隐间隔。

525

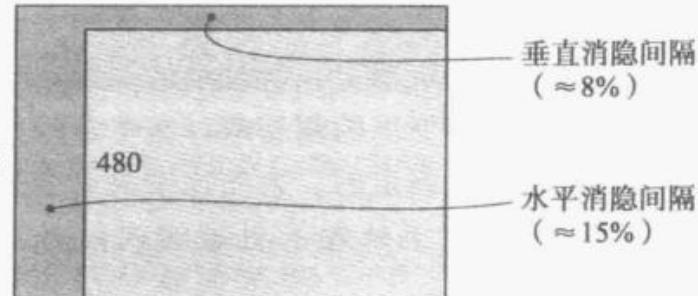


图 8-2 消隐间隔

用于 CRT 的水平和垂直消隐间隔占据行时间和帧时间的大部分：在标清视频（SD）中，垂直消隐大概消耗每个帧周期的 8%。在高清视频（HD）中，该部分被减少到大约 4%。

在一个视频接口，无论是模拟的还是数字的，同步信息（同步）是在消隐间隔期间传送。原则上数字视频接口只是传输伴随最低限度必要同步信息的有源像素。然而，数字视频接口使用的接口速率是与传统的模拟设备消隐间隔匹配的。额外的数据容量用来传送音频信号、字幕、测试信号、错误检测，或校正信息或其他数据或元数据。

8.3 扫描参数

在逐行扫描（或顺序）时，图像行扫描按顺序从顶部到底部以足以描绘运动的图像速率进行扫描。图 8-3 所示的顶部给出了四个基本扫描参数：

- 总线 (L_T) 包含所有的扫描线，包括垂直消隐间隔和图像扫描线。
- 图像具有包含图片的 N_R 图像行（原来称为活跃的行数， L_A ）。
- 总线样本 (S_{TL}) 包含在总线中包括水平消隐在内的采样间隔。
- 图像具有包含图片的数控图像列 N_C （通常某些活跃扫描线 (S_{AL}) 允许采取与消隐

电平不同的值)。

产品孔径如图 8-3 所示, 包括由 N_R 行 N_C 列组成的阵列。产品孔径内的采样点包括像素阵列, 它们都是动态的。所有其他采样间隔都含有消隐间隔, 它们是非动态的或“消隐”的。

所有标准 SD 和 HD 格式图像的 N_C 和 N_R 都是偶数。图像的水平中心位于两亮度样本中间, 并且画面的垂直中心位于垂直方向的两个图像行之间。

模拟信号中垂直消隐间隔通常携带如 VITS (垂直间隔测试信号)、VITC (垂直间隔时间码) 和隐藏式字幕垂直间隔信息。用户显示设备必须消隐这些行。数字视频接口中的垂直和水平消隐间隔可以用来传输类似于音频辅助 (ANC) 的数据。

图像阵列中的像素只有在获取、处理、存储和显示过程中被表示。然而, 某些处理操作 (如空间滤波) 只使用围绕受检像素的小邻域信息。在没有更好的可用信息情况下, 我们把图像阵列的像素置成黑色。在画面的左侧边沿, 如果最左侧像素的视频信号具有一个与黑色明显不同的值, 当信号传输时是通过模拟滤波器或数字滤波器处理, 将会产生一个称为振铃的后果。类似的情况也会出现在图片的右边沿。对于演播室内的视频, 信号呈现最大幅度, 或衰减到消隐电平, 理想情况下按升余弦包络经过多个传输采样点。

动态采样点不仅包括图片, 而且还包含传输采样点, 如图 8-4 所示。演播室设备应保持产品孔径所许可的最大图像范围, 同时还受制于消隐传送。

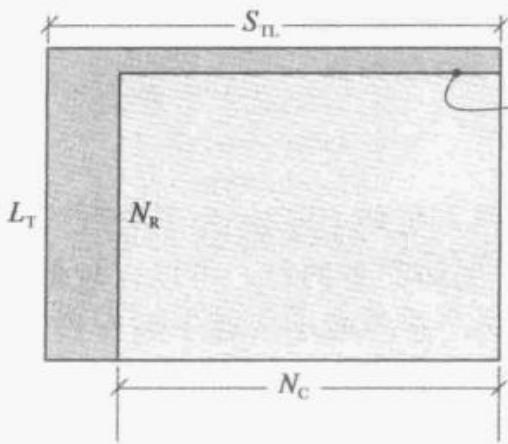


图 8-3 产品孔径

包括图像阵列, 即由 N_R 行组成的数控阵列。消隐间隔位于生产光圈之外。在这里, 消隐间隔是指深色阴影部分。每帧的有效像素数由 N_C 和 N_R 得出。采样率 (f_S) 是根据 S_{TL} 、 L_T 和帧频得出的。

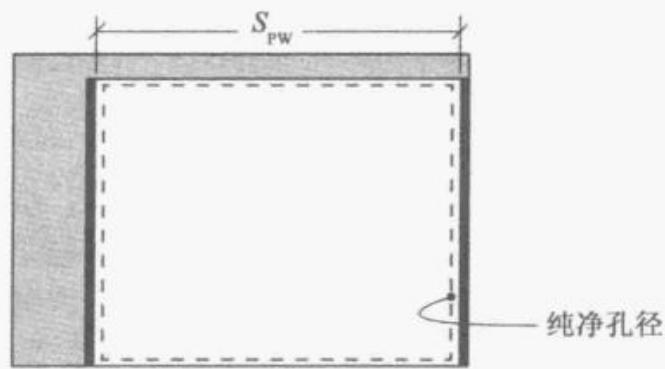


图 8-4 纯净孔径

纯净孔径应不受由过滤所产生的伪影的影响。纯净孔径不包括消隐传送样本, 消隐传送样本用图像宽度左边沿和右边沿外部的黑色条带表示, 由每个图像宽度 (S_{pw}) 样本数量定义。

将图像阵列当做像素阵列处理, 而不考虑各像素点光功率的空间分布, 比如像素点轮廓, 或更偏技术上的点扩散函数 (PSF)。如果点轮廓是这样的, 在相邻图像行 (扫描线) 的强度分布之间会有显著的间隙, 当观看者小于一定距离, 图像的结构将对于观察者是可见的。扫描线之间的间隙是图像行 (扫描线)、俯仰和点轮廓的函数。光斑大小是用在 50% 电量条件下光斑的直径来度量的。对于一个给定图像的行间距, 如果扫描线不可见, 较小的光斑尺寸将迫使观看者从距显示器更远的距离观察。

8.4 隔行扫描

隔行扫描在给定的观看距离, 闪烁的灵敏度和数据传输速率的情况下, 使得传统的 CRT 显示器静态空间分辨率相对于逐行扫描有所提高, 但同时会带来闪烁。除了垂直方

向上会留有一些间隙外，整个图像都会被扫描。然后 $(1/50)$ s 或 $(1/60)$ s 以后，整个图像高度再次进行扫描，但会向垂直方向偏移，以填补间隙。从而一帧图像包含两个场，分别用第一个和第二个表示。扫描机制如图 8-5 所示。传统上同一扫描标准也可运用在整个电视系统中，所以隔行扫描不仅用于显示，而且运用于整个系统，包括采样、记录、处理、分发和传输。



图 8-5 隔行扫描格式

隔行扫描格式取自两个场的完整图片（帧），每一个都包含图像的总行数的一半。第二个场与第一个场有半个帧时间的延迟。这个例子显示了 10 个图像行。在模拟扫描中，总扫描线数为奇数时会影响隔行扫描（例如 525, 625 或 1125）

在台式机和计算机中非隔行（逐行）扫描是非常普遍的；同时，逐行扫描已引用到高清数字电视中。然而，隔行扫描技术仍然普遍运用于标清和广泛用于广播、高清。隔行扫描到逐行扫描的转换（I-P），称为去隔行，是令人遗憾的，但却是必要的隔行扫描的副产品。

CRT 显示器现在已经过时。目前用于视频占主导地位的显示技术（如液晶和等离子面板）具有相对较长的工作周期，并且它们不闪烁。使用隔行扫描的理由不存在了。然而，隔行扫描仍然广泛使用。

8.5 抖动

视觉闪烁敏感性源于广域影响：一个存在闪烁的显示器，例如 CRT，为了克服闪烁需要图像的高度足够快速地闪现，小规模的画面细节，例如交替的行，可以以较低的速率进行传输。对于逐行扫描，扫描线的可视性限制了光斑尺寸的减小。对于隔行扫描，这个限制由于两个因素而放宽。但是，隔行引入了新的约束，即抖动。

如果一图像具有与该图像行间距相当的垂直细节，例如，如果对图 8-6 所示水平线对的精细图案进行扫描，那么隔行显示会使所述第一场和第二场的内容明显不同。在通常的帧率（50 或 60Hz），这样将导致抖动，通常是以小的闪烁或非常迅速的上下运动的形式表现出来。如果这样的图像信息占有相当大的面积，那么闪烁被感知，而不是仅仅是抖动。抖动有时称为行间闪烁，但闪烁定义为一个广域的效果，所以行间闪烁是一个不太好的现象。

抖动不仅仅产生于如图 8-6 所示的黑白混合线的退化图像，还产生于拥有高对比度的垂直细节的普通图像。高品质的视频摄像机内置用于衰减垂直细节的光学空间低通滤波器，缺失这部分将容易产生抖动。当计算机生成图像（CGI）采用隔行扫描时，垂直细节必须进行过滤，以避免闪烁。完成这一功能的信号处理器件称为抖动过滤器。



图 8-6 抖动

如果画面在指定的行间距下，用一个没有垂直滤波的摄像头进行扫描，然后显示在采用隔行扫描的短工作周期的显示器（如 CRT 显示器）上，这种情况下抖动就会产生抖动

8.6 模拟系统中的隔行扫描

在模拟视频中，隔行扫描是通过以一个恒定速率（常为 50 或 60Hz）垂直扫描，同时按照其一半速率的奇数倍进行水平扫描实现的。在北美和日本，标清场频为 59.94Hz，行频 (f_H) 为场频的 $525/2$ ($262\frac{1}{2}$) 倍。在亚洲、澳大利亚和欧洲，场频为 50Hz，行频为场频的 $625/2$ ($312\frac{1}{2}$) 倍。

图 8-7 表示出了水平驱动 (HD) 和垂直驱动 (VD) 的脉冲信号，为了在模拟设备中便于隔行扫描，这些信号会一次性分布在整个工作环境中。这些信号都会被组合同步（或复合同步）信号取代；垂直扫描是通过有 $2\frac{1}{2}$ 线或 3 线的总持续时间的宽脉冲触发。同步通常是施加到视频信号上的，以避免独立的分配电路。模拟同步编码在一个低于黑色基准的电平上。

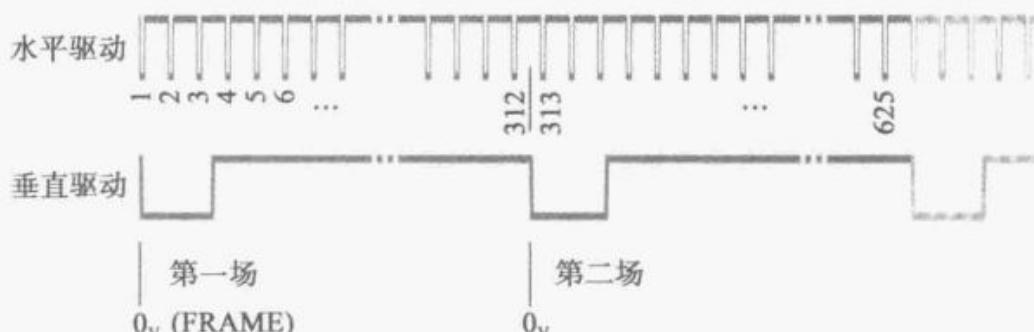


图 8-7 水平和垂直驱动脉冲

8.7 隔行扫描和逐行扫描

对于一个给定的观看距离，光点尺寸变小会改善清晰度。然而，如果光点尺寸的减小幅度超过一个限度，这个限度取决于显示器像素点的轮廓，扫描线和像素之一将变得可见，或将产生混叠现象。原则上，带宽或点轮廓的提高会减小潜在的观看距离，获得更大的画面视角。然而，由于用户对观看距离形成了期望值，我们假定一个恒定的观看距离并且分辨率有一定的改善。

逐行扫描和隔行扫描的粗略概念比较在图 8-8 中已给出。乍一看，隔行扫描系统提供 2 倍于像素的扫描速率，即分辨率的 2 倍，而逐行扫描系统采用相同的数据容量和相同的帧频。由于抖动，实际隔行扫描系统的空间分辨率并不是逐行扫描系统在相同的数据速率时空间分辨率的 2 倍。传统上摄像机的设计避免了产生太多的垂直细节，以及令人反感的抖动。然而，分辨率增加是通过隔行扫描这一历来就很被看重的因素来实现的。这些改善是以引入一些混淆和垂直运动伪影为代价的。此外，隔行扫描使得难以处理运动的图像序列，这将 8.9 节进行说明。

检查图 8-8 所示的隔行扫描（底部），想象图像元素以每场像素阵列一行的速率在图像上缓慢向下移动，对于 480i29.97 系统，在 $(1/60)$ s 内传输图像高度的 $1/480$ ，或在 8s 内传输全部图片高度。由于隔行扫描，一半图像的垂直信息将会丢失！以其他速率，图像中垂直细节的一些部分将丢失。采用隔行扫描，垂直运动可能会导致严重的运动伪影。用来避免这种现象的技术在行扫描中讨论。

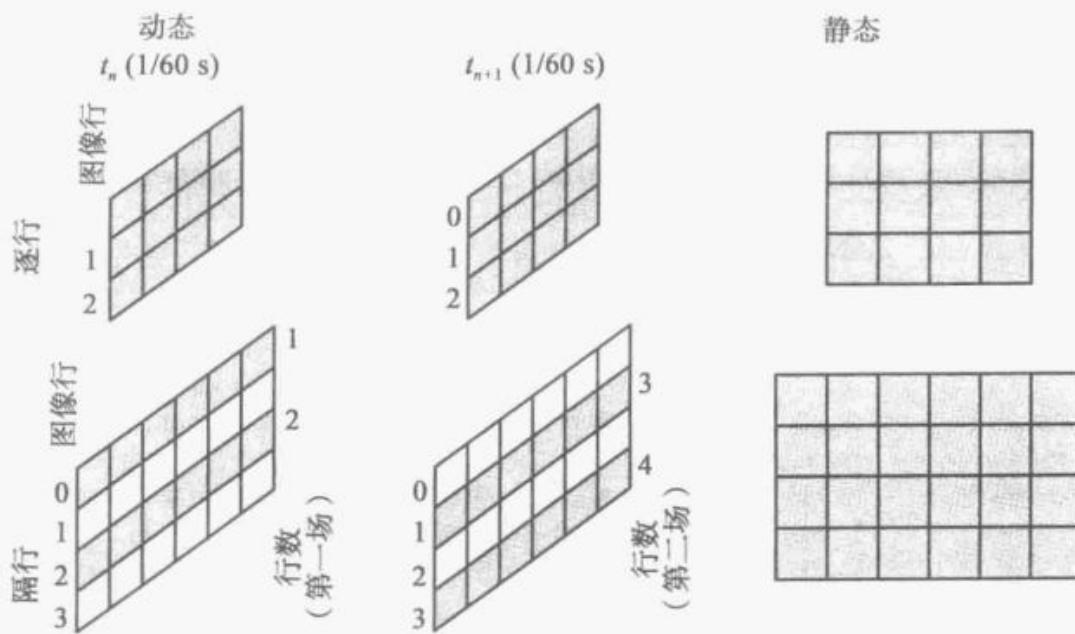


图 8-8 逐行扫描和隔行扫描

8.8 扫描表示法

在计算中，一种显示格式可以由一对数字来表示，即像素在跨越图像宽度的像素数——图像中的列数和跨越图像高度的像素数——图像的行数。方形采样是隐含的。这个符号并不表示刷新率（另外，显示格式可以象征性地表示为 VGA、SVGA、XGA 等，如表 8-2 所示。绝大多数计算机用户对这种表示法是不知道的）。

传统上视频扫描由每帧图像中扫描线的总数（图像行加上同步和垂直消隐的开销）、斜线和以赫兹为单位的场速率表示（斜线和 1:1 被追加到表示逐行扫描，除此之外，隔行扫描是隐含的。斜线和 2:1 使隔行扫描更加明确），对于标清，在北美和日本采用 525/59.94/2:1 扫描；在欧洲，亚洲和澳洲采用 625/50/2:1 扫描。高清视频（HD）出现之前，这些都是用于广播的唯一扫描系统。

数字技术产生了新的扫描和显示标准。传统的扫描符号不能充分表征新的扫描系统。一个新的表征符号出现了，如图 8-9 所示，扫描用图像的行数表示，跟在后面的 p 代表逐行扫描，i 代表隔行扫描，其次是帧频。传统的视频表示法（如 625/50）是不一致的，每帧中的并列线与每秒帧数一致。有些人特别是隔行扫描的支持者似乎是有意把这种混淆带到未来，把旧式 525/59.94 表示为 480i59.94。我强烈希望使用帧频。

在我用的符号中，传统的 525/59.94/2:1 视频记为 480i29.97；常规 625/50/2:1 视频记为 576i25。高清（HD）系统包括 720p60 和 1080i30。高清电影用 720p24 和 1080p24。在新符号中宽高比是不明确的：720p, 1080i 和 1080p 是隐含 16:9 的，因为对于这些系统没有 4:3 的标准，但 480i 或 480p 可能潜在有两种传统的 4:3 或 16:9 宽屏幅型比。

表 8-2 计算中的扫描

符 号	像素阵列/像素
VGA	640×480
SVGA	800×600
XGA	1024×768
WXGA	1366×768
SXGA	1280×1024
UXGA	1600×1200
WUXGA	1920×1200
QXGA	2048×1536



图 8-9 现代图像格式符号

8.9 运动描绘

讲述闪烁，刷新率和帧频的章节，我在选择刷新率时，列出了能感知到的因素。为了避免令人讨厌的闪烁，有必要用 CRT 显示器以比其描绘运动图像更高的速度闪现一张图像。根据图像质量要求和观看条件，不同的应用采用了不同的刷新率。刷新率通常设计成一个视频系统，一旦选择，就不能轻易地改变。

最小化闪烁可以通过使显示设备在大多数帧时间内产生稳定的非闪烁光或脉冲光实现。你可能认为一个非闪烁显示器比一个闪烁的显示器更加适合；许多现代设备不闪烁。然而，用一像素占空比接近 100% 的显示器，也就是说，工作时间接近帧时间——如果观看者的视线跟踪跨越图像移动的像素，该元素将被看作是被涂污的。随着例如高清的宽视角等因素使得眼睛跟踪速度增加，这个问题变得更加严重（对于图像采样和显示的时间特性的详细信息，请参阅我的 SMPTE 论文）。

通过 2-3 下拉，每秒 24 帧的电影转换为每秒 60 场的隔行扫描视频。第一电影帧被转移到两个视频场，然后第二电影帧被转移到三个视频场；周期重复。2-3 下拉方式通常用来制作视频，在 59.94Hz，不是 60Hz 时，电影的运行速度比每秒 24 帧的速度慢 0.1%。该方法见第 34 章介绍。2-3 下拉技术可以应用到以每秒 59.94 帧或 60 帧的逐行扫描视频。

每秒 24 帧的电影使用 2-2 下拉转换到 576i 的视频：每个电影帧被扫描到两个视频场（或帧）；电影运行速度比捕获速率快 4%。该过程在 2-3 下拉内容中更全面地进行了描述。

8.10 分段帧 (24PsF)

渐进式分段帧方案有时用在以每秒 24 帧处理图像的高清设备中。该过程表示为 24PsF，采用逐行扫描：整个图像采样一次，为减少抖动，进行垂直滤波既是不必要的也是不可取的。然而，扫描线会重新排列，以满足工作室和录音分配的交错顺序。该计划提供了隔行扫描处理和记录设备的一些兼容性。

8.11 视频系统的分类

电视广播开始时，没有足够的信道容量可用来发送三个独立的颜色分量。NTSC 和 PAL 技术被设计成将三个颜色成分结合成一个单一的复合信号（编码）。这在用户场所的复合视频中仍然在使用。然而，无论是 $Y'P_BP_R$ 模拟元件或 $Y'C_BC_R$ 数字组件，现代视频设备，包括所有的消费类数字视频设备，以及所有高清设备，都使用分量视频。

视频系统可以划分为高清分量、标清分量、复合标清。独立地，一个系统可以分为模拟系统或数字系统。按照相关联的颜色编码方案，表 8-3 给出了 6 种分类。复合 NTSC 和 PAL 视频编码只用在 480i 和 576i 的系统，高清系统只使用分量视频。S-视频是模拟分量视频和复合模拟 NTSC 或 PAL 的混合。在表 8-3 中 S-视频被归类在自己的第七个（混合）类别。

表 8-3 视频系统分类

		模 拟	数 字
高清	分量	$R'G'B'$, 709 $Y'P_BP_R$	$4:2:2$ 709 $Y'C_BC_R$
	分量	$R'G'B'$, 601 $Y'P_BP_R$	$4:2:2$ 601 $Y'C_BC_R$
	混合	S-video	
	复合	NTSC, PAL	$4f_{sc}$ (过时的)

8.12 系统之间的转换

在视频中，编码传统上指的是将一组 $R'G'B'$ 分量转换成 NTSC 或 PAL 复合信号。编码可以从 $R'G'B'$ ， $Y'C_BC_R$ ，或 $Y'P_BP_R$ 分量开始，或者可能涉及从 $R'G'B'$ 开始的矩阵变

换以形成亮度信号 (Y') 和中间体 $[U, V]$ 分量；然后正交调制形成调制的色度 (C)；之后得出亮度和色度。解码传统上指的是从 NTSC 或 PAL 复合信号到 $R'G'B'$ 的转换。解码包括亮度/色度分离，正交解调恢复 $[U, V]$ ，然后扩展到恢复 $[C_B, C_R]$ 或 $[P_B, P_R]$ ，或亮度和色度来恢复 $R'G'B$ 的矩阵化。现在的编码和译码已成为如 JPEG、M-JPEG、MPEG 等一般术语或其他编码或解码过程。

转码传统上指的是在具有相同的扫描标准而不同的颜色编码方法之间的转换。分量视频转码涉及色度插值，矩阵变换和色度亚采样。复合视频转码涉及解码，然后重新编码为其他标准。随着压缩存储和数字分销的出现，转码这个词现在正在应用在重新编码压缩的位流，或解压缩然后重新压缩等各个领域。

扫描转换是指在具有不同空间结构的扫描标准之间转换，并且不对时间处理。在消费类视频和桌面视频，扫描转换通常称为缩放。

如果输入和输出的帧频不同，则操作认为是涉及帧频转换。运动的描绘容易受损。

传统上，上变频是指从标清到高清的转换，下变频是指从高清到标清的转换。这些术语所指的是在同一帧频的信号作为输入的转换；现在帧频转换可能包含其中。高品质上变频和下变频需要空间插值。那个过程最好是在一个逐行扫描格式中进行：如果源视频是隔行扫描的，即使目标格式是隔行扫描，中间去隔行扫描是必需的。

标准转换是传统名词，意指在具有不同的帧频扫描标准之间转换。这个词暗示类似的像素数（如 480i 和 576i 之间转换），但现在标准转换可能会包括上变频或下变频。标准的转换需要一个场存储器或帧存储器；为实现高品质，它需要几个场存储器和运动补偿内插。对广播公司和消费者来讲，在不同帧频之间的标准转换的复杂性使得欧洲的元器件难以在北美和日本使用，反之亦然。

第9章

分 辨 率

为了避免图像显示可见的像素结构，相邻的显示元件产生一些重叠光分布是可取的，详见第7章。此外，为了避免在图像捕获中的空间混叠，在跨越相邻传感器元件的灵敏度分布中一些重叠是必要的。这种重叠降低了锐度，但对连续色调的图像是有益的。在本章中，我将解释与清晰度密切相关的分辨率。

分辨率是一个多重模糊的术语，而正确的是指空间现象。我们所指采样按照8位分辨率很让人混乱，所以使用精度或量化代替。在计算中，通常使用术语“分辨率”来指定图像矩阵的列数和行数，也就是要表达的像素数。这种使用无视信号处理效果。像素要优先考虑分辨率数就很难讨论其代表或展现给观众的影像细节。本章将详细介绍分辨率，但首先要介绍的幅频响应和带宽的概念。

9.1 幅频响应和带宽

为了表征小单元的采样、处理或显示，而不是分析确定（小）尺寸的一个单元，我们分析一组紧密间隔的相同的单元，表征单元的间距。这允许使用数学变换方法，特别是傅里叶变换和 z 变换。

图9-1所示顶部曲线给出了一维正弦波从零频率“扫描”到高频的测试信号（这可能是关于时间的一维函数，如音频波形，或图像的一个扫描线的亮度波形）。一个典型的光学或电子成像系统涉及时间或空间的传输，这导致系统的响应在高频率处被削弱，如图9-1所示的中间曲线。该波形的包络线即为系统的幅频响应，如图9-1的底部所示，电气工程师可以简单地称其为频率响应。限定词幅度使得它与其他如相位频率响应等频率功能相区别。

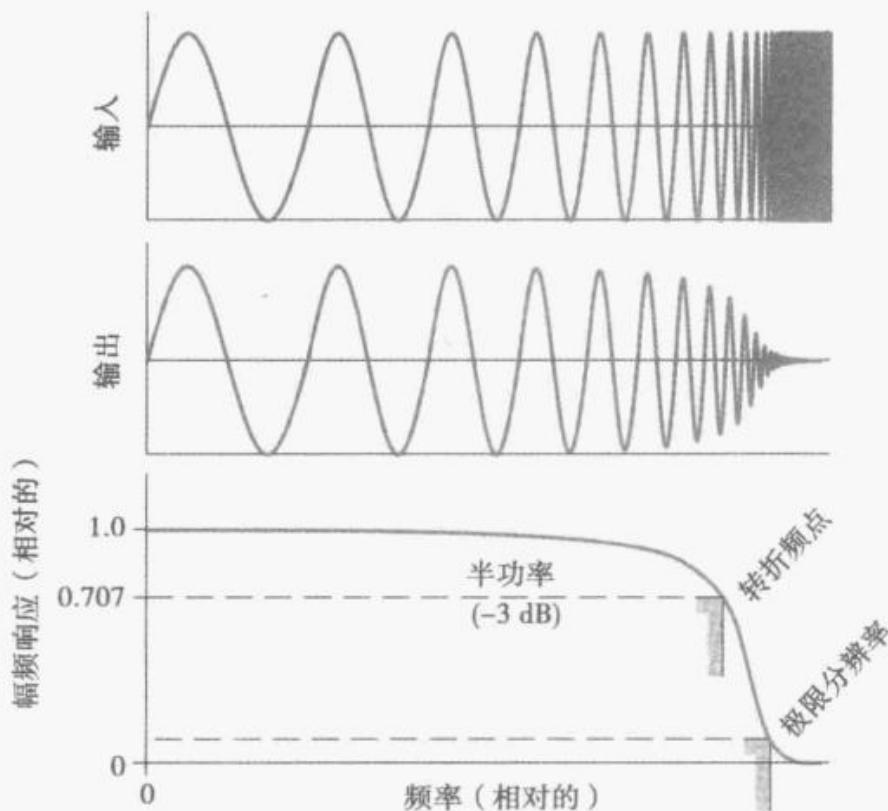


图9-1 幅频响应曲线

带宽表示一个系统可以捕捉、记录、处理或传输的频率范围。半功率带宽（也称为 -3dB 带宽）是在其信号幅度下降 3dB 处测量的，即在基准频率（通常是零频率的值，或DC）下峰值的 0.707 。有用的视觉信息通常在高于带宽频率处。在影像学中，极限分辨率是由视觉确定的。

在黑白成像系统中一个模拟或数字电子信号可以改变状态的最大速率受频率响应限制，因此能用带宽表示。

图 9-1 显示了抽象的输入和输出信号。当讨论光学系统的带宽时，带宽大小与强度成比例是隐含的。当讨论视频信号的带宽时，输入和输出的电信号是经伽马校正的，这也是隐含的。

许多数字技术人员用带宽代指数据速率，但是这两个术语有不同的概念。带宽指的是模拟或数字信号内容的频率范围。数据速率是指数字传输能力，独立于任何潜在的信号内容。一个典型的演播室高清亮度信号具有 30MHz 信号带宽和 74.25MB/s 的数据传输速率，这两个词显然是不能互换的。

9.2 视觉灵敏度

当验光师测量视力时使用类似于图 9-2 所示的图表。这个测试结果取决于观看距离。该测试的标准观测距离是 20ft 。在这个距离下，字母轮廓对应着 $20/20$ 行中的六十分之一度 $((1/60)^\circ$ ， 1 弧分)。这大致是正常视力的角度识别的限制。

可视角度可以用遵从天文学家的拇指描绘规则法进行估算得出，如图 9-3 所示。当伸出手臂，拇指的关节对着约 2° 。全手掌对着 10° 左右，小指的指甲对着大约 1° （满月的角弦大约是 0.5° ）。



图 9-2 视力表

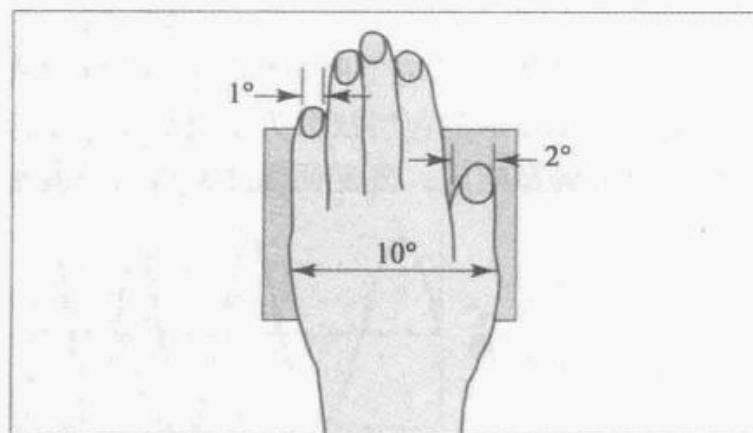


图 9-3 天文学家的拇指规则法

图 9-3 所示的天文学家的拇指规则法可以粗略测量对向角。伸直手臂，手掌对着大约 10° 的视角。这里，手掌覆盖的幅型比为 $4:3$ 的矩形。如果该矩形是标清图像，此时观看者将大致位于最佳观看距离。

9.3 观看距离和角度

显示一个白平场到一具有典型像素间距的显示器上，如果观看者是位于比在所显示表面相邻图像行（扫描线）对着的距离还近 1 弧分 $((1/60)^\circ)$ 或更近，像素结构很可能是可见的。

为实现观看像素间距对应 $(1/60)^\circ$ ，观看距离应为图像的行间距 (d) 的 3400 倍，也就是说， 3400 除以像素密度。例如，像素密度 (ppi) ：

$$\text{距离} \approx 3400 \cdot d \approx \frac{3400}{\text{ppi}}; 3400 \approx \frac{1}{\sin\left(\frac{1}{60}\right)} \quad (9-1)$$

因此，在3400倍的图像行的距离处，每度约有60个像素。观看距离用数字表示为画面高度的倍数，然后用3400除以图像行(N_R)：

$$\text{距离} \approx \frac{3400}{L_A} \text{PH} \quad (9-2)$$

标清图像有大约480图像行(图片线)。在大约7倍图片的高度(PH)的距离处图像行对着($1/60$)°，如图9-4所示，整个画面大约600个像素。图片角度大约是11°，如图9-5所示。伸出手臂，手掌应该刚好盖住图片的宽度。这个距离约是显示器对角线的4.25倍，如图9-6所示的空白处。对于1080像素图像行的高清，能够产生的($1/60$)°像素对边的观看距离约为3.2 PH(见图9-4的底部)时，大约为显示器对角线的1.5倍。

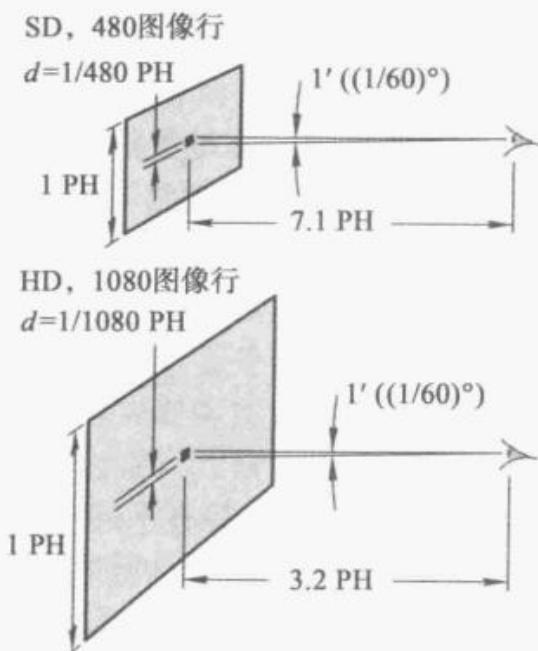


图9-4 观看距离

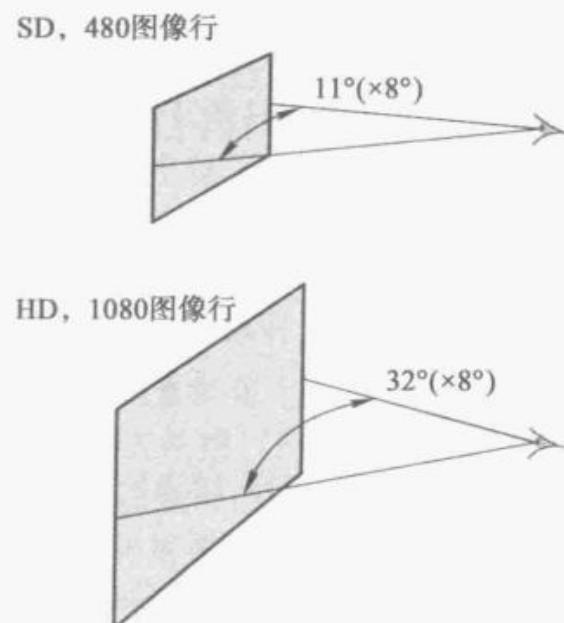


图9-5 SD的画角

图9-6所示的4:3幅型比的图像高度为对角线的 $3/5$ ，最佳观看距离为传统的视频对角线的4.25倍。16:9幅型比的画面高度约为对角线的一半，2百万像素高清的最佳观看距离为对角线的1.5倍。

对于标清，在那样的观察距离下总水平画角约为11°。观看者倾向于选择像素结构不可见的观看距离，一个像素的角弦因此保留。因此，较高像素数的主要作用是提供宽画角。如图9-5所示，对于 1920×1080 HD，水平可视角度3倍于SD的11°而到33°。高清晰度并不是将6倍的像素数挤到同一个视角里！相反，整个图像可以占有观察者的视野的一个更大的区域。这个话题会在可视距离中进一步讨论。

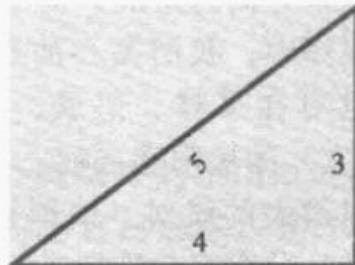


图9-6 4:3幅型比的图像高度

9.4 克尔效应

最大分辨率不能在早期电视系统中传输，并寄希望于奈奎斯特的工作。1934年，克尔发表了一篇量化最大理论分辨率的论文并通过了RCA的实验电视系统。他称这个数为 k ，也就是后来的克尔因子(很少记为 K)。克尔的第一篇论文给出了因子0.64，但他没能完整描述相应的实验方法。随后一篇论文(1940年)详述了实验方法，并给出了在不同

条件下的因子为 0.8。

克尔的 k 因子是由主观的、不客观的标准确定的。如果所测试的系统有一个点轮廓类似于高斯分布，当其间距减小超过一定值时，测试图上的紧密间隔的线将不再处理。如果被测照相机有一个不同寻常小的点，或显示器有一个尖锐的分布（如箱分布），则 k 是由随着间隔减小而引入的令人不快的伪影来确定，同时这也是一个主观标准。

克尔和其他作者发表了辩解各种数值因素的理论推导；Hsu 发表了全面的评论。在我看来，这样的数值措施是如此定义不清和不可靠的，现在它们是没用的。Hsu 说：

克尔因子定义是如此含糊，研究人员已经名正言顺地用不同的理论和实验技术来获得不同的 k 值。

今天我认为量化数字克尔因子是低级的科学。然而，雷·克尔向电视科学做了重要贡献，我认为用克尔效应来纪念他很合适。

在视频系统——包括图像采样、信号处理、传输和显示——克尔效应是指相对于奈奎斯特极限，由光功率引起的空间色散，从而导致分辨率损失。一些分散是必要的，以避免捕获后的混叠；也避免一个显示错误的扫描线或像素结构。

克尔 1934 年的论文只涉及逐行扫描。随着隔行扫描系统出现，由过量的垂直细节引起的抖动变得很明显。为了减小抖动到可接受的水平，有必要大幅度降低垂直分辨率，使得其低于精心设计的具有相同光斑尺寸的给定 k 的逐行扫描系统，而隔行扫描系统具有相同的光斑大小必须有较低的 k 。很多人都把这个因素归因到“克尔因子”，但研究人员如 Mitsuhashi 只是定义这种减小为隔行因子。

9.5 分辨率

720×480 像素的 SD， 1280×720 像素和 1920×1080 像素的 HD 都有不同的像素数。传送特定数目像素的图像的质量取决于图像数据的特性（例如，数据是否为栅格锁定或奈奎斯特滤波），也取决于显示设备的特性（例如，它是否有框或高斯重建的性质）。

然而，在计算中分辨率这个词仅仅指的是像素阵列中的垂直和水平像素的数量，当然这是在不考虑任何捕捉重叠或显示重叠的情况下，可能减少的图像中的细节。一个系统可以被描述为具有 1152×864 像素的分辨率，该系统共拥有约一百万像素（百万像素，或 1Mpx）。可以这样理解，“分辨率”不依赖于单个像素在显示屏上是否被识别出。

数字图像系统的分辨率是由图像的宽度和高度的像素数界定的。然而，由于图像细节在频率方面的增加，信号处理和光学效应引起相应减少，即使是在采样所规定的范围内。在视频中，我们关心的是展现给观众的分辨率，同时我们也对在采样、记录、处理和显示过程中的频率响应（带宽）的极限感兴趣。在视频中，分辨率涉及显示器上可以消失的线对（或周期）的最大数目。这是一个主观标准！分辨率与感知灵敏度有关。

分辨率通常以空间频率表示，单位是水平上每图像宽度周期数（C/PW）和垂直方向上每图像高度周期数（C/PH），或单位与这些方面密切相关。图 9-7 描述了一个分辨率测试图。在呈现的方位，它扫过水平频率，并且可以用来估计水平分辨率。转动 90°，它可以通过垂直频率扫描，并由此估计垂直分辨率。

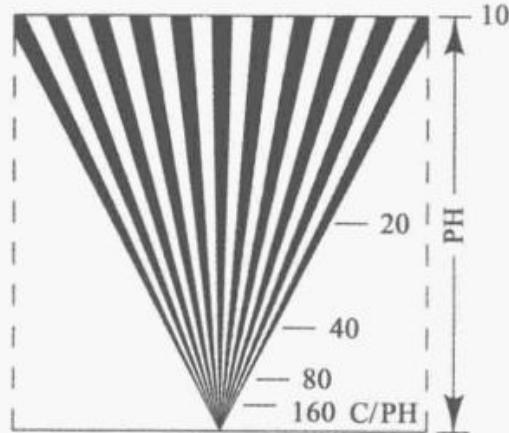


图 9-7 分辨率楔形图案

9.6 视频中的分辨率

在图像传感器或显示装置中空间现象可能限制了垂直分辨率和水平分辨率。模拟处理中，记录和传输限制了带宽，从而只影响水平分辨率。传统上视频中的分辨率指的就是水平分辨率。

TVL/PH 中的分辨率，通常称为“电视线”，是垂直黑白线对（周期）的 2 倍，此时，它是在等同于图片高度的水平距离上被识别的。

在用户设备垂直重采样已变得普遍；重采样可能会影响垂直分辨率。在基于变换的压缩（例如 JPEG，DV 和 MPEG）中，与像素之间重叠相当的色散会发生，这会影响水平分辨率和垂直分辨率。

9.7 观看距离

在 SD 和 HD 中像素数是由相应的图像格式确定的。9.3 节解释了最佳观看距离是在扫描线间距对着约 $1/60^\circ$ 处。如果采样的图像在比那个距离更近时观看，扫描线或像素是比较容易可见的。用典型的显示器，SD 是适合在大约 7 倍 PH 处观看；1080iHD 适合于在约 3 倍 PH 值处或更接近的距离处观看。

计算机用户往往坐在扫描线间距对着的角度大于 $1/60^\circ$ 处 也许只是那个距离的一半。然而，在这样近的距离，单个像素可能是明显的，甚至可能是令人反感的，并且连续色调图像的质量肯定会受到影响。

最接近观看距离是受像素数限制的，但是，图像中像素的可见性或扫描线结构取决于许多其他因素，如照相机的 MTF（调制传递函数），光斑轮廓（PSF）和频率响应。原则上，如果任何这些因素降低了图像中的细节质量，则最佳观看距离就变远。但是，消费用户已经形成了一种期望，SD 视频最好是在大约 7 倍 PH 值处看，随着人们逐渐熟悉 HD 视频，他们会认为最好是在大约 3 倍 PH 值处观看 HD 视频。

反补贴论点是，基于用户客厅的尺寸。未发表的研究中，伯尼·莱希纳发现，北美洲的观众倾向于在约 9ft 处看 SD 电视。在英国类似的实验，理查德·杰克逊发现他们趋向在 3m 处。观看距离有时也称为 Lechner 距离，在欧洲称为杰克逊距离！这些数字与式 (9-2) 保持一致，应用到 SD 显示器采用了 27in (70cm) 的对角线。

与其说带宽或点轮廓的提高使观看距离减小，因此获得更大的画角，我们假设观看距离是固定的，然后就说分辨率得到了改善。

9.8 再谈隔行扫描

现在我们可以再谈一下隔行扫描的参数。用典型的照度和环绕条件的家用电视接收机，50Hz 或 60Hz 的垂直扫描速率足以克服闪烁。正如在实际的垂直扫描率，有可能在交替隔行垂直扫描中闪现图像不会产生闪烁。这就是隔行扫描。由于视觉系统的时间灵敏度在高空间频率减小，该方案是可能的。

然而通过垂直细节尺度接近扫描线间距，会引入抖动。抖动可以通过减少垂直细节降低到可接受的水平，可能为原来的 0.7 倍。就其本身而言，这种在垂直细节的减少将使观看距离推回至逐行扫描的 1.4 倍。

然而，在给定数据容量情况下，为了保持与隔行扫描系统相同的清晰度，所有其他条件都相同，隔行扫描中只有一半的图像数据需要在每个垂直扫描周期（场）进行传输。对于给定的帧频，这样每次扫描数据的减少使每帧像素数加倍。

获得的像素可能由以下三种方式之一被利用：通过行数加倍，通过列数加倍，或通过按图像的行和列的比例分配额外的像素。采取第三种方法，加倍的像素数可以相等地在水平和垂直上分布，行数和列数都增加到原来的 1.4 倍。因而观看距离可以降低到逐行扫描时的 7/10，重新赢得由抖动引起所损失的观看距离，并产生与逐行扫描相当的性能。

理论上，由于隔行扫描追加的像素不应该相等地分布到这两个维度。相反，图像的行数应增加为原来的 1.4×1.2 （即 1.7）倍和图像的列数约为原来的 1.2 倍。因子 1.4 的增加缓解了抖动。逐行扫描系统中行数和列数剩下的 1.2 倍增加使得观看距离适度而显著缩短，同时也会改善画角。

抖动和扫描线的可见性是与图像的行数成反比的，这是一维的数量限制。然而，清晰度是与像素数成正比的，这是二维（面）的数量限制。为了克服在同样画角下的抖动，1.4 倍的图像行是必需的，但是，1.2 倍的图像行和 1.2 倍图像列仍可以提高画角。

在半个世纪前的早期电视中采用隔行扫描，而没有采用逐行扫描。其他条件一样，如数据速率、帧频、光斑尺寸和观看距离等，隔行扫描的各种优势已经显示出来了。

- 忽视抖动的引入，仅仅考虑静态像素阵列，对于一个给定的带宽和帧频隔行扫描提供了两倍的静态分辨率。
- 如果你考虑到与逐行扫描图像相同大小的隔行扫描图像，并在相同的距离处观察，也就是说，保持画角，扫描线可见度会降低。

第10章 恒定照度

视频系统使用一个近似明度的组件以及表示色彩和缺少明度的两个分量来传送彩色图像数据。第26章将详细介绍正比于光功率的线性RGB的加权和如何形成照度的。色彩专家使用恒定照度是指这个和是恒定的。发送相对照度能够被重构的单个分量是基于恒定照度的原理确定的。最好是一个非线性传递函数作用于该分量来施以感知统一编码。

标准的视频系统并不严格遵守这一原则，但是它们实现了工程近似。色彩专家的线性RGB加权并不是计算得来的。相反，非线性传递函数被单独地应用到每个线状光的RGB分量中，然后非线性伽马校正 $R'G'B'$ 分量的加权和形成我所说的亮度（许多视频工程师也称为数量照度）。在标准的视频系统，亮度是利用色彩学的理论RGB加权因子进行编码的，但在方框图中亮度不同于色彩专家所期望的那样：在视频中，伽马校正的应用早于矩阵校正，这并不是色彩专家所期望的。

传统上，灰度单分量传输增强了黑白电视的兼容性。相比照度，人眼视觉对颜色敏感度不高。配置“黑白”信息为一个分量，强化了色度亚采样，以利用视觉对色度的低敏锐度来减小其他两个分量的数据传输速率（传统上即为带宽）。在彩色成像中，尽管黑白兼容性没有要求，编码“黑白”分量还是很合理的（例如，在JPEG格式）。

我把“黑白”放在引号中。在电视中，所传输的信号表示灰度，而不是只有黑和白，那么，现在灰度将是一个更好的词。

传统视频指的是“代表照度的信号”或“照度信号”或“照度分量”。所有这些词曾经都是有道理的，但是它们现在是危险的；如果要使用“照度”，这表明相对照度(Y)可以从该分量中进行解码。然而，如果没有严格遵守恒定照度的原理，照度不能单从灰度分量解码；其他两个分量（通常是 C_B 和 C_R ）是必要的。

本章将阐述目前所有的视频系统为什么以及如何从恒定照度的原理脱离出来。如果你愿意接受这种理论，那么你可以放心地略过本章，并继续学习亮度和色度内容，同时我也将介绍亮度和色差信号如何形成和完成亚采样的。

10.1 恒定照度原理

理想情况下，彩色视频中所谓单色分量会模仿灰度系统：色彩学原理之RGB和CIE XYZ之间的转换，相对照度将被计算为（线性光）的R、G和B三刺激值合理的加权总和，如图10-1所示。在解码器中，逆矩阵将重建线性的R、G和B三刺激值，如图10-2所示。

计算两个色差（色度）分量来增强色度亚采样，这些将通过独立的通道转达给解码器。

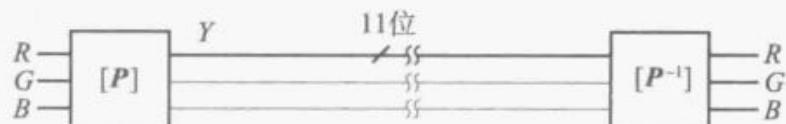


图10-1 相对照度的形成

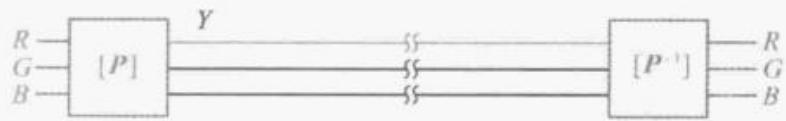


图10-2 假设的色度分量（线性光）

忽略色度分量现在：不管如何处理他们，在一个真正的恒定照度系统中所有的相对照度是可从灰度分量恢复出来的。

如果直接传输相对照度，11位或更多是必要的，如图 10-1 所示。如果我们使用非线性图像编码，8位勉强够用来增强感知均匀性，我们可以令相对照度服从非线性传递函数来模仿视觉亮度的敏感性。亮度可以近似为 CIE L^* ， L^* 大致是相对照度的 0.42-幂函数，如图 10-3 所示。

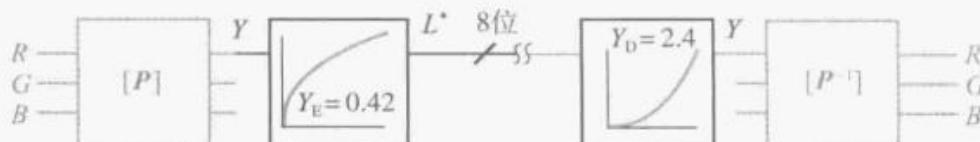


图 10-3 对非线性编码的相对照度进行编码

该解码器将采用逆传递函数，如图 10-4 所示

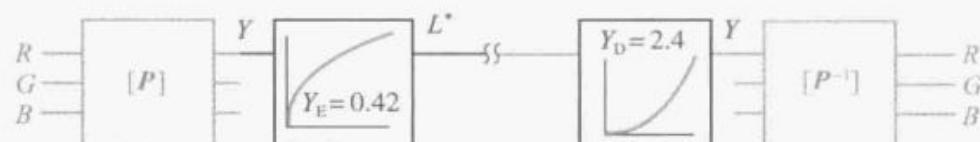


图 10-4 对非线性编码的相对照度进行解码

如果视频系统都以这种方式进行操作，这将符合恒定照度的原理：所有的相对照度会存在于并可追溯到灰度分量。

10.2 CRT 补偿

对理论框图来说是不幸的，但对视频来说却是好事，稍后即将看到。传统 CRT 显示器的电子枪引入了底数大约为 2.4 的幂指数函数，如图 10-5 所示。

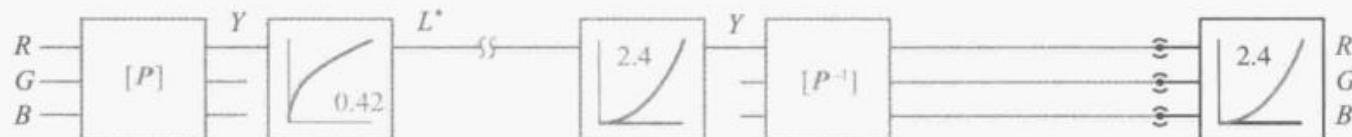


图 10-5 CRT 传递函数

在一个恒定照度的系统中，译码器将不得不颠倒显示器的幂函数。这将需要在 CRT 的前面插入一个补偿传递函数（大约为以 1/2.4 为底的幂函数），如图 10-6 所示。

解码器现在将包括两个幂函数：一个是指数接近 2.4 逆 L^* 函数，用来反转感知统一编码，另一个是指数为 1/2.4 的幂函数，即约 0.42，以补偿 CRT 的非线性。图 10-6 表示一个理想化的，真正的恒定照度视频系统的结构图。

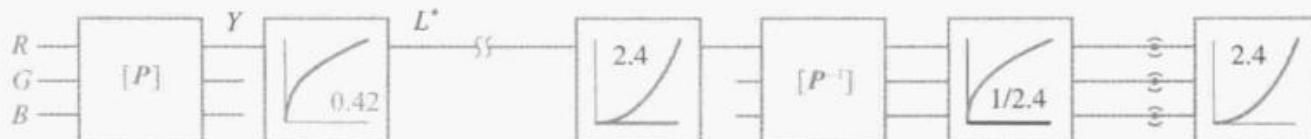


图 10-6 CRT 传递函数的补偿

10.3 从恒定照度的分离

每个解码器采用两个非线性传递函数既昂贵又不切实际。

注意幂函数的指数是 2.4 和 1/2.4，这两个函数是互逆的！为了避免将两个幂函数集

成到有一个解码器的电子设备的复杂性，我们通过重新排列框图，开始交换矩阵和 CRT 补偿的“操作顺序”，如图 10-7 所示。

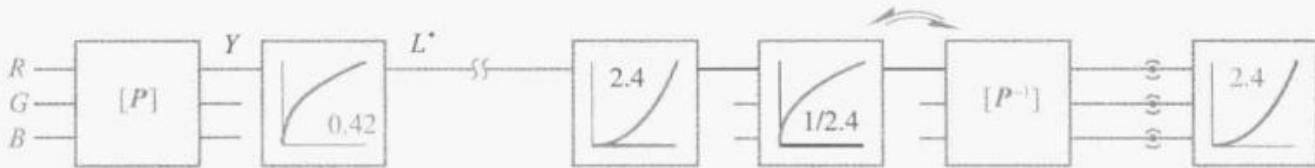


图 10-7 重排译码器

重排后，两个幂函数是相邻的。由于这两个函数是互逆的，这二者的结合没有任何实际效果。这两个函数都可从解码器移除，如图 10-8 所示。



图 10-8 简化译码器

解码器的信号处理只简单地反转编码矩阵。以 2.4 为指数的幂函数是 CRT 显示器内在的，替代显示技术呈现出与从信号值到三色类似的映射。

重排解码器需要编码器也重新排列，以便反映解码器和实现正确的端到端的原始 RGB 三刺激值的再现。

图 10-9 表示所有视频系统的基本信号流，将会在后面的章节中加以阐述。



图 10-9 重排编码器

10.4 亮度

图 10-9 所示的重新排列流程图，数学上不等价于图 10-1 到图 10-4 所示的组合！在图 10-9 中，编码器矩阵不对（线性光）三刺激值光信号进行处理，也不计算相对照度。相反，以 Y' 表示的非线性亮度被计算和传输。亮度涉及工程近似：这个系统不再严格遵守恒定照度原则（尽管它经常被错误地认为这样做）。

没有学过色彩学的电视工程师经常误将亮度 (Y' , luma) 说为照度 (luminance)，并表示为 Y ，这导致了很大的混乱。

在重排列的编码器中，我们不再使用 CIE L^* 来优化感知均匀性；相反，我们使用 CRT 固有的传递函数的逆运算。以 0.42 指数的幂函数来近似完成感知统一编码，并且再现正比于原始场景的三色的值。

接下来图片渲染一章将阐明指数 0.42 必须在一个正常的场景进行修改至约 0.5，以适应一个感知的效果。这种改变出于美观、显示，以及观看条件的考虑。理想情况下，显示系统应该根据显示器和观察条件对图片渲染有所调整，但却很少这样做！

10.5 照度转换成色度的“泄露”

到现在为止，我们已经忽略了色差分量。在图 10-9 所示的重新排列框图中，色差分量（见图 10-10）是非线性（伽马校正） $R'G'B'$ 的“矩阵式”。

在一个真正的恒定照度系统，无论如何处理色差信号，所有的相对照度是由灰度分量携带。在重新排列的系统中，大多数的相对照度通过 Y' 信道被传送。然而，在某种程度上 Y' 不等于 Y ，一些相对照度被认为是转换到色差分量过程中的“泄漏”。如果色差分量没有亚采样，例如，在一个 $Y'C_B C_R$, 4 : 4 : 4 的系统，该泄漏将是无关紧要的。

然而，为强化亚采样，色差分量均精确地成形！所以，我们现在把注意力放在亚采样上。

图 10-11 表示了图 10-10 所示色度路径上增加了亚采样滤波器的实际框图。以传统的编码方式，一些相对照度要横穿色度通道。图 10-12 上方显示了在色条上红绿过渡的以 4 : 2 : 2 色度亚采样的理想化 $Y'C_B C_R$ 波形。图 10-13 实线所示的是传统的解码之后的照度。亚采样不仅消除了颜色分量中的细节，还消除了“泄漏”的相对照度的细节。我们要问，“到底丢失了什么？”如图 10-14 空白处所示，出现在标准视频红绿色条的暗带中的理论偏差还是挺明显的。

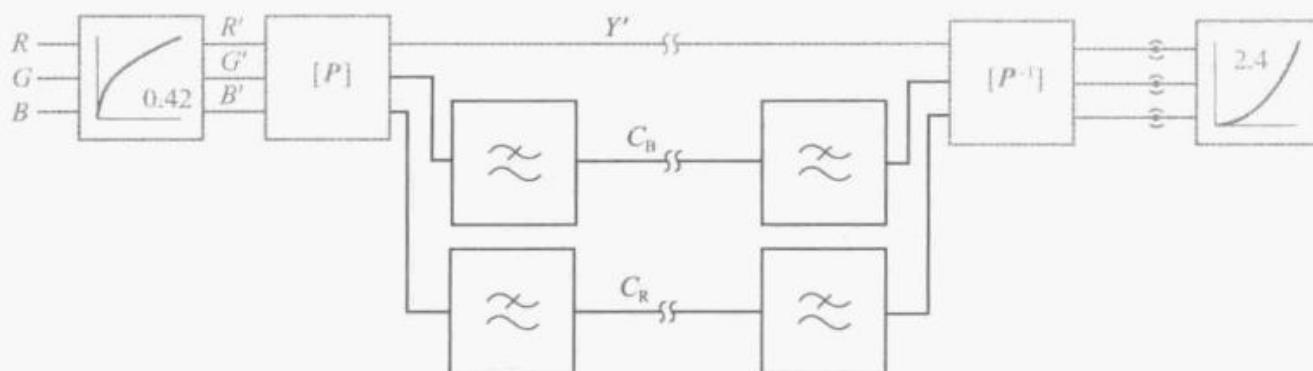


图 10-10 色度分量

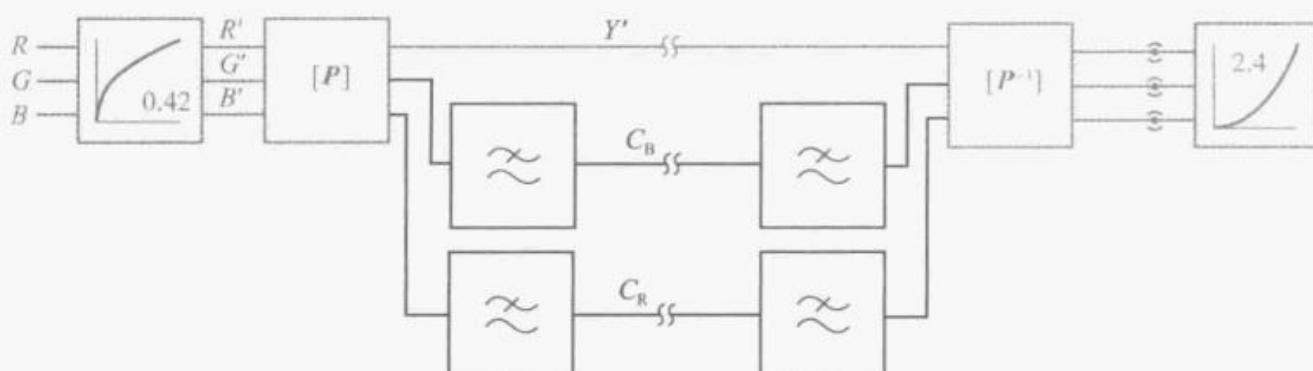


图 10-11 二次采样的色度分量

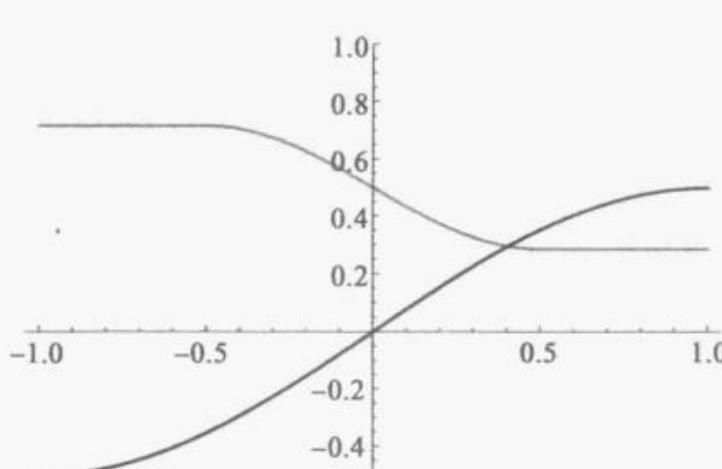
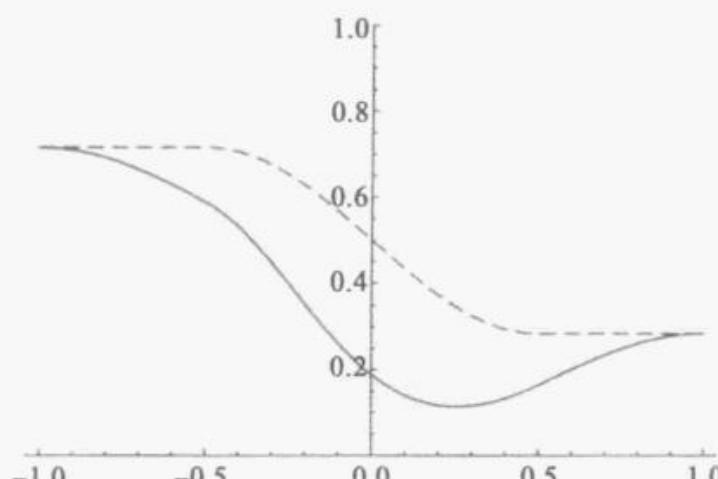
图 10-12 SD 色条中绿色-洋红色过渡段 Y' 和 C_B/C_R 的波形

图 10-13 色条中的绿色-洋红色过渡段照度波形

采用传统的视频编码，在照度细节存在于饱和色彩的区域，相对照度被错误地重建：相对照度被重建得太暗，饱和度降低。高频照度的不准确传输是不严格遵守恒定照度原则所必须付出的代价。在正常的图像中，这样的“利文斯顿”错误只有专家才能察觉到，但他们很少被发觉，更不用说反感。

总结视频系统中的信号编码：首先，类似于平方根的非线性传递函数，即伽马校正，应用到每个线性的 R 、 G 和 B 三刺激值，以形成 R' 、 G' 和 B' 分量。然后，将非线性分量适当加权和计算以形成亮度信号 (Y')。亮度逼近视觉的明度响应。色差分量蓝色减去亮度 ($B' - Y'$) 和红也减去亮度 ($R' - Y'$) 从而形成（亮度， $B' - Y'$ 和 $R' - Y'$ 可以从 R' 、 G' 和 B' 通过一个 3×3 矩阵同时计算）。然后，色差分量二次采样（过滤），可采用几种方案，包括 $4:2:2$ 、 $4:1:1$ 和 $4:2:0$ 。

本章概述了在 NTSC 制式发展过程中，恒定照度的工程近似如何一步步被采纳，而不是“真正的”恒定照度。该工程定制运用非常好，已经运用到分量视频系统（SD 和 HD）和现代压缩系统，如 JPEG、MPEG 和 H.264。

自 2000 年以来，大多数的电视接收机已经融合了数字信号处理，并且消除了 1950 年所做的工程参数：图 10-6 所示的两个非线性函数现在很容易通过查找表（LUT）来实现。一些较真的人认为，在现代社会，我们应该取消近似而采用正确的理论方法。然而，图 10-9 所示的为 SD 和 HD 的视频设施框架（图 10-11 所示的为增加了色度亚采样部分）。改变视频的框图并处以巨大的转换负担看起来似乎不合理，除非可以证明有足够的优点。我很欣赏理论论证，但是我不清楚这样的变化是否会带来显著的好处，所以我认为我们不应该改变视频的框图。

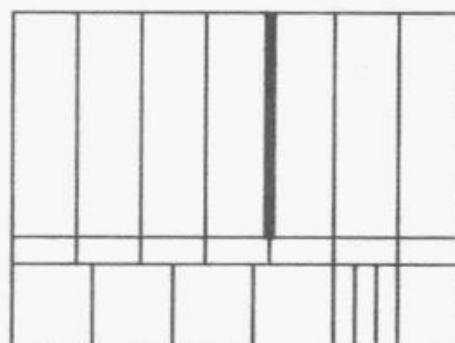


图 10-14 不遵守恒定照度

第11章

图片渲染

在一个阳光明媚的天气中午观看花园里的花朵。日落半小时后再看同样的花园。在物理上，花的谱图都没有变，只是被扩展到较低的照度水平。然而，花在日落之后色彩明显变淡：颜色饱和度随照度降低而减小。这就是亨特效果，首次（1952年）由著名色彩学家R.W.G.亨特描述得出。再现的图像通常以它们被捕获照度的一小部分观察，或许 $1/100$ 或 $1/1000$ 。如果重建照度与场景照度成正比，重现图像就会显得不那么丰富多彩，而且比原来的场景有较低的对比度。

要重现与原始场景类似的对比度和颜色饱和度，我们必须改变图像的特征。工程师或物理学家努力实现成像系统中的数学线性关系，然而，所要求的改变会导致再现的照度线性偏移。难点在于：我们可以实现数学线性关系或者实现正确的外观，但是我们不能同时一举两得！成功的商业成像系统以牺牲数学线性关系来实现正确的感知结果。

如果在观看环境的“白色”的照度明显小于在其被捕获的环境的“白色”，图像的色调范围必须改变。另外一个修正的原因是环绕效果，这也是我正要解释的。

11.1 环绕效果

人的视觉适应极其广泛的观看条件，这点在23.1节详述。当感兴趣的区域被明亮的元素包围，一种涉及适应的机制增加了我们对小明亮度变化的灵敏度。直观地说，来自四周明亮区域的光可以看作是溢出或散射进入我们视野的各个领域，包括感兴趣区域，减小了其明显的对比度。粗略地讲，视觉系统通过“拉伸”其对比度范围补偿这种影响，以增加处于明亮环绕中较深区域存在的可见度。相反，当感兴趣区域被相对暗色包围，视觉系统的对比度范围减小：我们辨别场景暗色元素的能力下降。该效果表现在图11-1中（来源于DeMarsh and Giorgianni）。环绕效果源于称为并发对比效应的感知现象，也称为侧抑制。

环绕效果具有图像中暗区的显示含义，如电影院的放映仪，35mm幻灯片放映或在客厅观看电视。如果图像以正确的相对照度再现，然后在黑暗或暗淡的环绕观看时，它会表现得缺乏对比度。

图像再现不是简单地与物理、数学、化学、电子等有关：感性因素起着至关重要的作用。

11.2 色调比例调整

色调比例调整是必要的，主要有两个理由：再现的照度通常比原始场景的照度大幅降

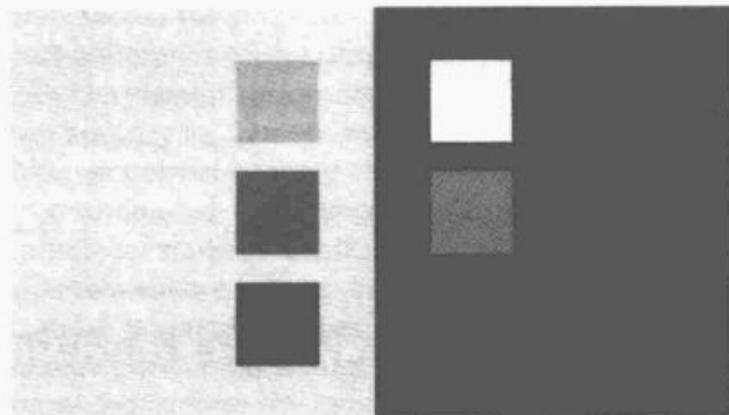


图11-1 环绕效果

低，还有就是再现图像的环绕很少比得上原始场景的环绕。另外两个因素促成色调比例调整，那就是对比度和镜面高光限制。

一个典型的原始场景具有大于 1000 : 1 的照度等级之比。然而，所捕获的图像的对比度是由摄像机光学耀斑限制的。显示器对比度一般受到物理因素和大约 1000 : 1 显示器耀斑的限制。

漫反射白指的是场景中发生漫反射的白色表面的照度。纸面普遍发生反射，白纸反射约 90% 的入射光，所以白色卡片近似漫反射的白色。然而，大多数场景包含了定向反射的发光对象。当在某些方向观看时，这些对象反射其照度可能 10 倍于弥漫性白色的镜面高光。再现装置中，仅仅使用显示器最大照度的 10%，我们很少能够重现弥漫性白色，从而不能完全准确地再现亮点的照度水平！也没有任何必要精确地重现亮点：令人信服的图像可以由将真正照度值打些折扣的高照度形成。为了有效地利用那些在图像显示系统的照度范围，亮点必须压缩。

11.3 渲染合并

我所提到的校正可通过使照度服从一个指数介于约 1.1 和 1.5 之间的端到端幂函数来实现，或者在彩色系统的情况下通过三刺激值实现。该指数主要取决于场景的照度与再现照度的比值，在一定程度上它取决于显示器物理和观看环境。几乎所有的图像再现系统需要一些色调范围调整。

第 10 章概述了视频中非线性编码。回忆的草图序列，图 11-2 表明了典型电视观看的校正可以通过在解码器加一个以 1.2 为指数的幂函数来实现。

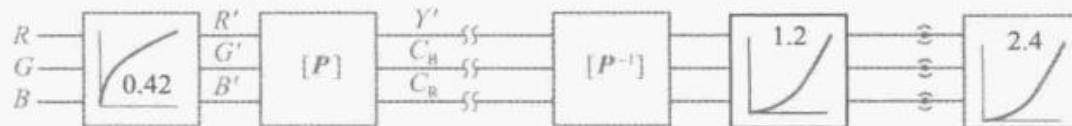


图 11-2 在解码器中施加图片渲染

注意到，一个幂函数已经是编码器的必要组成部分。不改变解码器，我们修改编码器的幂函数来使指数接近为 0.5，而不是物理上的精确值 0.42。

串联编码器以 0.5 为指数的幂函数和解码器中以 2.4 为指数的幂函数产生在昏暗环绕的电视显示所需的端到端的以 1.2 为指数的幂函数。为了恢复场景的三刺激值，编码传递函数应该简单地被反转；解码函数则近似于以 2.0 为指数的幂函数，如图 11-3 所示。

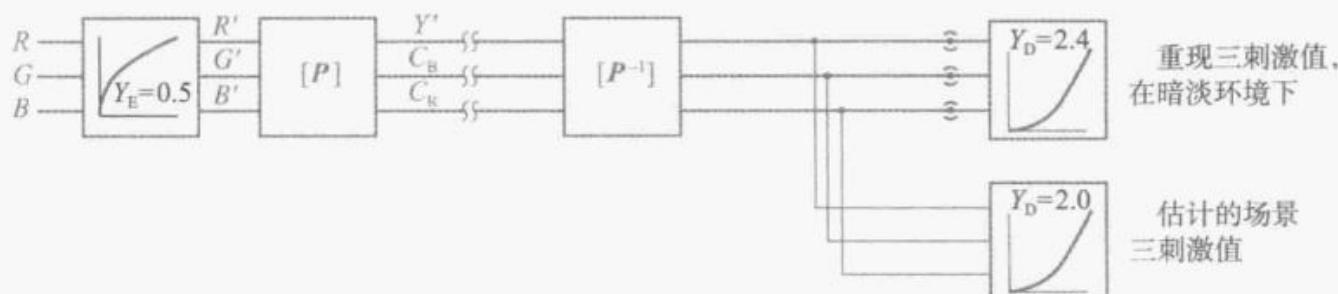


图 11-3 在编码器中施加图片渲染

CRT 的有效幂函数的指数变化取决于亮度控制的设置。在黑暗的观看环境，如家庭影院，显示器的亮度设置可以低至将解码器的幂函数有效指数提高到 2.6，端到端幂函数指数约为 1.3。在明亮的环境如在桌面环境中的计算机，亮度将增加，有效指数将降低至约 2.2，从而减小端到端指数至约 1.1。

电影、电视和办公 CRT 的编码指数、解码指数和端到端幂函数已在表 11-1 中列出。

表 11-1 成像系统端到端的幂函数

图像系统	编码指数	“广告”指数	解码指数	典型周围环境	对比度	端到端指数
影院（电影投影）	0.6	0.6	2.5	黑（0%）	100 : 1	1.5
高清，演播室控制（BT. 709/BT. 1886）	0.5	0.45	2.4	很暗（1%）	1000 : 1	1.2
高清，客厅（典型）	0.5	0.45	2.4	暗（5%）	400 : 1	1.2
办公室（sRGB 标准，典型）	0.45	0.42	2.2	平均（20%）	100 : 1	1.1

在胶片系统中，必要的校正通常都会设计到摄影底片和打印薄膜的传递函数中。投影的图像通常用于黑暗环境中；端到端幂函数已作出调整，指数比串联编码器和解码器的大得多，通常约为 1.5，从而使得场景的对比度范围在显示时变大。在电影胶片中，校正是通过相机底片的传递函数（“伽马”值约为 0.6）和印刷底片的传递函数（“伽马”值约为 2.5）的组合实现的。

我所描述的视频系统是假设它们使用一个单纯的 0.5 幂律编码函数。考虑实际因素，纯幂函数的修改是通过在黑色区域处插入线性段实现的，这一点我将在第 27 章中解释。BT. 709 标准指数为 0.45（建议），但是，线性段的插入，曲线的纯幂函数段的偏移和缩放，使得约 0.51 为指数的幂函数得以最好地对整体曲线进行描述（在这种情况下将伽马值定为 0.45 是一种误导）。

11.4 桌面计算中的图像渲染

在桌面计算机环境中，周围环境尤其明亮，环绕的亮度比典型电视观看环境更亮。此时要求视频的端到端的指数小于 1.2，大概为 1.1 是合适的。然而，台式计算机用于各种不同的观看条件下。将不同格式的图像优化到几个较好观看条件下是不实际的！这时需要选择一个特定编码函数。在不同的观看条件下实现最佳再现，需要在显示时间上选择一个合适的修正。从技术上讲，这很容易实现：现代计算机显示子系统具有硬件查找表（LUT），能够动态地加载相应的曲线。然而，培训用户做出合适的选择是一个挑战。这就需要传感器来检测环境光和相应的算法，以实现适当的修正（主要是通过改变显示器的伽马值）。这些计划已经实现商业化，但目前还没有标准。

正当开发用于桌面计算的 sRGB 标准，局部的必然性，依赖于观看的修正，并没得到认可。该标准规定以单纯 2.2 为指数的幂函数进行解码，但标准还描述了用黑色附近的线性段和有效指数大约为 0.45 的编码标准是什么。细读 sRGB 标准证实，sRGB 标准是显示参考；用线性段定义的视频是从屏幕表面的三刺激色值的映射成的 sRGB 代码值。sRGB “编码”函数无法与 BT. 709 的参考光电转换函数 OECF 相比。显示器的 sRGB 标准元器件应做到纯 2.2 为指数的幂函数，没有任何直线段。

在 sRGB 标准下，指数（建议）为 1/2.4（约 0.42）。然而，线性段的插入，曲线的纯幂函数段的偏移量和尺度，使得约 0.45 指数的幂函数最好地描述了整体的曲线，见 27.7 节。

摄像机、胶片相机、电影摄影机和数码相机都是从现实世界中捕捉图像的。当原始场景或物体的图像被捕获时，引入渲染是重要的。然而，在桌面计算机中使用的扫描仪很少扫描原始对象，它们通常扫描如摄影照片或印刷图像等复制品。当这些复制品被扫描时，第一成像处理中的渲染已经进行。调整原来的渲染是应该的，但引入适合于扫描真实场景或对象的渲染是不明智的。

第 12 章

亮度和色度简介

视频系统中传送图像数据的形式是，一个分量表示明度，两个分量表示颜色而忽略明度。相比照度敏感度，该方案采用降低视觉色敏感度的方法：只要传送明度的全部细节，颜色分量中的细节在亚采样过程中会减小，即滤波（平均滤波）。本章介绍亮度和色度编码的概念，细节将在第 28 章详述。

12.1 亮度

数字成像系统存在一定量的噪声是不可避免的。如 1.5 节所述，编码的目的就是使得噪声具有从黑到白的整个色调范围感知类似的效果。明度分量是以感知均匀性的方式进行传送的，这样能最大限度地减小处理、记录和传输过程中引入的噪声（或量化误差）。

理想情况下，噪声会通过形成与 CIE 照度成比例的信号被最小化，结果会是线性 R、G、B 三刺激值适当加权和。然后，该信号将服从感知均匀性的传递函数，详细的介绍参见 24.5 节函数。如第 10 章所介绍的，视频中有实际要考虑的因素，这要求我们以相反的顺序执行这些操作。首先，非线性传递函数，即伽马校正应用到每个线性 R、G、B 三色信号上：我们施加类似于平方根的传递函数，并大致相当于 CIE 明度 (L^*) 函数。然后得到非线性的 R' 、 G' 和 B' 分量的加权和以形成代表明度的亮度信号 (Y')。SD 将使用标准化的 BT. 601 因子为：

$$^{601}Y' = 0.299 R' + 0.587 G' + 0.114 B' \quad (12-1)$$

然而，HD 亮度编码不同于 SD 亮度编码！BT. 709 标准规定了这些因子为：

$$^{709}Y' = 0.2126 R' + 0.7152 G' + 0.0722 B' \quad (12-2)$$

12.2 照度简述

照度和符号 Y 是由色彩学标准组织 CIE 在约 75 年前制定的。然而，在视频中，尽管视频信号分量已经服从一非线性传递函数，照度是指具有代表性的视频信号的照度。在视频兴起时，非线性信号用 Y' 表示，其中的首要符号表示非线性处理。但在过去 50 年中首要符号始终未出现，而今天，照度和相应的记号 Y 都与 CIE 定义相冲突，从而产生了模糊概念！这导致了很大的混乱，如计算机图形学教材和 YIQ 或 YUV 颜色空间方面的数字图像处理教科书中的错误陈述，即 Y 分量与 CIE 照度是相同的！

本书的照度是按 CIE 定义的：亮度指的是视频信号；我很仔细地用质数来指定非线性量的值。然而，许多视频工程师、计算机图形学从业人员和图像处理专家胡乱使用这些术语。你必须仔细地确定被应用到这个词和符号中的解释是线性的还是非线性的。

12.3 色差编码（色度）

在分量视频中，传递彩色信息的 3 分量是分别传送的。 $R'G'B'$ 信号并不是被直接传输的，根据色彩信息，利用相对较弱的色彩敏锐度减小数据容量，同时保持充分的亮度细节。首先，亮度值是根据式 (12-1)（或式 (12-2)）得到的。然后，基于伽马校正两个色

差信号， B' 减去亮度和 R' 减去亮度，即 $B'-Y'$ 和 $R'-Y'$ ，由矩阵变换得到。最后，亚采样（滤波）在色差（或色度）分量方面减少细节信息，这点我会再做介绍。亚采样使得在任何合理的观看距离处都没有清晰度损失。

亮度和色差可以由 R' 、 G' 和 B' 通过一个 3×3 的矩阵乘法来计算。

12.3.1 $Y'P_B P_R$

在模拟分量视频中， $B'-Y'$ 和 $R'-Y'$ 缩放形成色差信号并表示为 P_B 和 P_R ，这两个信号将被模拟为低通滤波（水平）到大约一半的亮度带宽。 Y' 、 P_B 和 P_R 各自具有统一偏移（即 $0\sim 1$ ， ± 0.5 ， ± 0.5 ）。在计算中，这些分量通常是表示 U 和 V 。

12.3.2 $Y'C_B C_R$

在数字分量视频（包括M-JPEG，MPEG和H.264标准）中， $B'-Y'$ 和 $R'-Y'$ 缩放成 C_B 和 C_R 分量，然后可以通过标记为 $4:2:2$ 或 $4:2:0$ 的数字滤波进行亚采样；之后将介绍采样。

12.3.3 $Y'UV$

在复合NTSC或PAL视频中， $B'-Y'$ 和 $R'-Y'$ 缩减成 U 和 V 分量。然后 U 和 V 进行低通滤波，并合成一个调制的色度分量（ C ）。然后亮度与调制的色度加和产生复合的NTSC或PAL信号。缩放 U 和 V 的大小使得所述复合信号 $(Y'+C)$ 的变化范围被限制到亮度统一偏移的 $-1/3$ 至 $4/3$ 。传统的 U 和 V 分量现在不再用了，现在用 UV 表示 $Y'P_B P_R$ （见上文）。

12.3.4 $Y'IQ$

在1953年复合NTSC视频进行了标准化，基本上是将 U 和 V 组件旋转 33° 并且轴交换得到 I 和 Q 分量。多余的细节应该从 Q 分量中除去，从而改善颜色质量。该计划从未取得显著搭配； I 和 Q 分量已经过时几十年了。

图12-1所示亚采样中， $2\times 2 R'G'B'$ 像素阵列经矩阵运算成一个亮度分量 Y' 和两个色差分量 C_B 和 C_R 。在对 C_B 和 C_R 进行亚采样过程中减少了色彩细节；如果完全保持亮度细节，就不会有退化。在该图中，将样本加上阴影用来表示它们的空间位置和范围。在MPEG-2标准中使用 $4:2:2$ ， $4:1:1$ 和 $4:2:0$ 格式， C_B 和 C_R 处于同一位置上（位于与亮度样本水平一致的位置上）。使用在JPEG/JFIF、H.261和MPEG-1标准中的 $4:2:0$ 格式， C_B 和 C_R 之间相隔一个间隙（处于采样样本之间）。广泛用于用户中的是576i DV的 $4:2:0$ 格式（不在这里讲解）， C_B 和 C_R 在每一帧内以行交替的方式相互垂直（从在左上角亮度样本的一个 C_R 样本开始）。

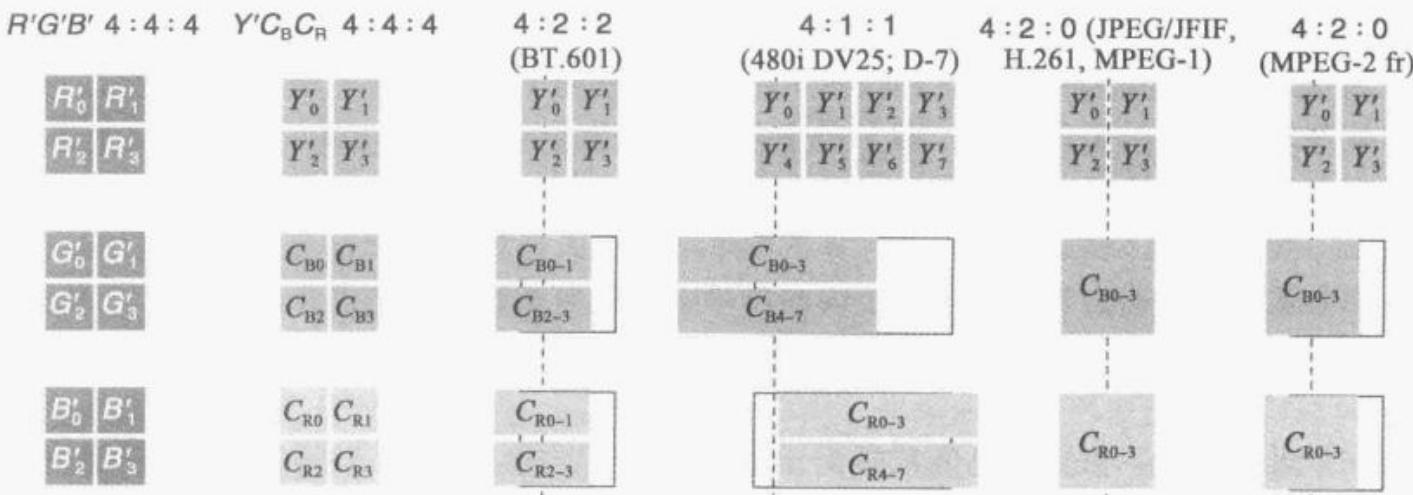


图12-1 色度亚采样

12.4 色度亚采样

12.4.1 4:4:4

在图 12-1 所示的亚采样中，左栏勾画一个 2×2 的 $R'G'B'$ 像素阵列（把这些 2×2 阵列想象为覆盖在显示表面上）。在亚采样之前，该采样被表示为 $R'G'B'4:4:4$ 。采用 8 位的采样率，这 2×2 的 $R'G'B'$ 阵列将消耗 12 字节。每个 $R'G'B'$ 三元组（像素）可以无损变换（“矩阵运算”）成 $Y'C_B C_R$ ，如第二栏所示，表示为 $Y'C_B C_R4:4:4$ 。

12.4.2 4:2:2

在数字分量视频中，使用三种方案中的一种对 C_B 和 C_R 进行亚采样会减小数据容量。按照 BT. 601 标准 $4:2:2$ ， $Y'C_B C_R$ 演播室数字视频采用 $4:2:2$ 采样：水平上 C_B 和 C_R 分量分别以为 2 因子进行亚采样。 C_B 和 C_R 一起进行采样，与偶数亮度样本保持一致。该 12 字节的 $R'G'B'$ 被减小到 8 字节，结果为 $1.5:1$ 的有损压缩。

12.4.3 4:1:1

某些数字视频系统，如 480i29.97 DV25，使用 $4:1:1$ 采样，由此 C_B 和 C_R 分量分别以 4 为因子进行水平亚采样，并与每第 4 个亮度样本处于同一位置。该 12 字节的 $R'G'B'$ 被减小到 6 字节，使得结果为 $2:1$ 的有损压缩。

12.4.4 4:2:0

这个方案用在 JPEG/JFIF、MPEG-2 和 H.264 标准中。 C_B 和 C_R 分别以 2 为因子进行水平亚采样和纵向采样。该 12 字节的 $R'G'B'$ 被减小到 6 字节， C_B 和 C_R 有效地垂直居中在图像行之间。 $4:2:0$ 的 3 个变量具有不同的垂直和水平的位置。在 MPEG-2 标准中， C_B 和 C_R 在水平上处于同一位置。在 JPEG/JFIF、H.261 和 MPEG-1 标准中， C_B 和 C_R 被填隙地放置在亮度样本之间。在 $4:2:0$ 的 DV 中， C_B 和 C_R 交替间隔。图 12-2 所示的总结了各种方案。

亚采样产生了 $1.5:1$ 或者 $2:1$ 的有损压缩效果，然而，在视频术语中，亚采样视频指的是未压缩：压缩这个词指的是 JPEG、M-JPEG、MPEG 或者 H.264 这些使用变换编码（离散余弦变换或者小波）的技术。

12.5 色度亚采样符号

在数字视频起初，亚采样符号是符合逻辑的；然而，技术的进步超过了符号表达的含义。如图 12-3 所示，我尽量说清现在所用的术语。除了外观，符号并未指定所占比例！第一位数最初指定亮度采样率为 $3\frac{3}{8}$ MHz；前面的数字是现在使用的采样率。初始位通常 是 4 位，这是因为所有的普通色度比是 2 的低次幂，如 4、2 或 1 ($3:1:1$ 亚采样已在 HD 生产系统（索尼 HDCAM）中商业化，3 作为首位数字。HDCAM SR 已取代 HDCAM，它采用 $4:2:2$ 或 $4:4:4$ 亚采样)。有些人用 $4:0:0$ 来表示灰阶（单色）。

图 12-2 所示的亚采样方案总结如下，当处于同一位置时，C 表示一个 (C_B , C_R) 样本对。否则（如在 DV 方案中）单个 C_B 和 C_R 符号表示相应的色度样本的中心。 Y' 表示亮度样本的中心。左列方案是逐行扫描的。右列方案是隔行扫描的。另外，实字母表示顶部区域的样本和阴影字母表示底场样本。

图 12-3 所示的色度亚采样符号表示（第一位数）水平亮度采样参考。第二位数指定相对亮度的 C_B 和 C_R 水平亚采样。第三位数最初指定的是 C_R 的水平亚采样。预测垂直亚采

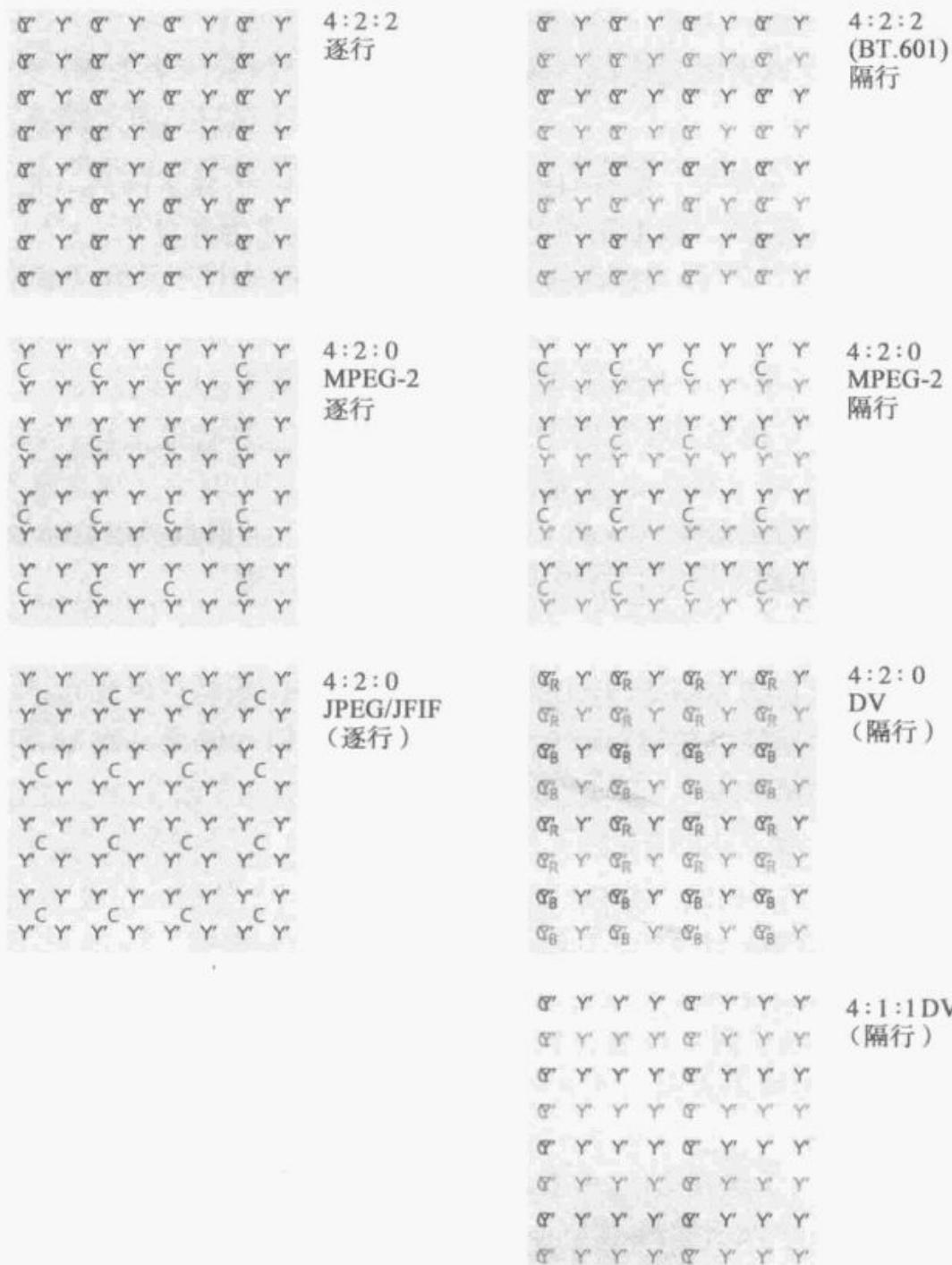


图 12-2 亚采样方案总结

样并未随符号的发展而发展；第三位数是零，现在表示 C_B 和 C_R 的 2:1 垂直下采样。

12.6 色度亚采样滤波器

在色度亚采样中，编码器经滤波后丢弃所选色差采样。解码器通过插值补充丢失样本。

要以最小的计算量进行 4:2:0 采样，有些系统只是在 2×2 阵列中对 C_B 和 C_R 取平均，如图 12-4 所示。要在转换回 $R'G'B'$ 之前插入丢失的色差样本值，低端系统简单复制 2×2 阵列中亚采样的 C_B 和 C_R 的值。这种技术广泛用于计算机中的 JPEG/JFIF 静止帧、M-JPEG、H.261 和 MPEG-1。这个简单的平均化处理使得亚采样色度在两个亮度样本之间产生一个有效的位置，我称之为间隙选

亮度水平采样参考（最初，亮度 f_s 是 $3\frac{3}{8}$ MHz 的多倍）
 C_B 和 C_R 水平因子（相对于第一个数）
 与第二个数一样，或者零，表明 C_B 和 C_R 是 2:1 垂直亚采样

4:2:2:4 如果存在，与亮度数一样，表明 alpha (关键) 分量

图 12-3 色度亚采样符号

址，与演播室视频的共存位置标准不一样。

一个按 BT. 601 标准要求的 4 : 2 : 2 水平亚采样的简单方法是使用上述 [1/4, 1/2, 1/4] 的权重，如图 12-5 所示。与隔行扫描相互影响相比，4 : 2 : 2 亚采样更具优势。

同位置的水平滤波器可与 [1/2, 1/2] 垂直平均相结合，如图 12-6 所示，可实现 MPEG-2 的 4 : 2 : 0 亚采样。

类似于图 12-4~图 12-6 所示的简单平均滤波器对静止帧有较好效果，其中所产生的任何混叠分量保持静止，或者用于桌面质量级视频。然而，在运动图像中，通过差的滤波引入的混叠分量易于以与相关联的场景元素不同的速率移动，并由此产生一个非常令人讨厌的伪影。高端数字视频设备使用复杂亚采样滤波器，其中 4 : 2 : 2 标准下的亚采样 $2 \times 1 C_B$ 和 C_R 对（或 4 : 2 : 0 标准下的 2×2 阵列）从周围几个样本中取权值。滤波器的权重、频率响应和过滤性能的关系在第 20 章中详细介绍，这些因子实现了一个适合 4 : 2 : 2 亚采样的高品质的 FIR 滤波器：[−1, 3, −6, 12, −24, 80, 128, 80, −24, 12, −6, 3, −1]/256。

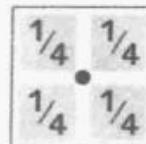


图 12-4 JPEG/JFIF 的间
隙性色度滤波器



图 12-5 同一位置 BT. 601 色度滤波器

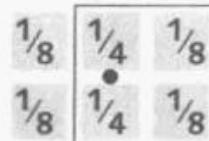


图 12-6 MPEG-2 4:2:0 标准下同位置色度滤波器

12.7 复合 NTSC 和 PAL 的色度

这里介绍的色差分量 $P_B P_R$ 和 $C_B C_R$ 通常称为色度分量。它们与亮度一起出现在分量视频系统中。同时还介绍了 UV 和 IQ 分量，这些中间变量在调制色度的形成过程中产生。

传统上，采用分别传输 3 个颜色分量来适应信道容量的不足。NTSC 技术设计将 3 个颜色分量结合成一个复合信号；PAL 技术是 NTSC 技术的细化和对 NTSC 技术到 576i 扫描的调整形成的（在 SECAM 中，3 个颜色分量也组合成一个信号。SECAM 是复合视频的一种形式，但该技术与 NTSC 和 PAL 相差甚远，并且现在它很少用在商业上）。

传统上 NTSC 和 PAL 编码器从 $R'G'B'$ 分量开始。在复合视频的鼎盛时期，数字编码器从 $Y'C_B C_R$ 分量开始。NTSC 或 PAL 编码包括以下步骤：

- 分量信号做矩阵变换和微调，从而形成色差信号 U 和 V（或 I 和 Q）。
- U 和 V（或 I 和 Q）经低通滤波，然后进行正交调制，使两个色差信号施加到一个未调制的彩色副载波上，以产生调制的色度信号（C）。
- 亮度和色度加和。在演播室视频中，加和采用频率交织的原则。

复合 NTSC 和 PAL 信号在传统上是模拟的。20 世纪 90 年代，数字复合（ $4f_{sc}$ ）系统开始使用，不过 $4f_{sc}$ 方案现在已经过时。8.11 节提到的，在普通用户和工业应用领域中分量视频已经替代了复合视频。欲了解更多信息，可参见第 14 章。

第13章

SD 分量介绍

第8章介绍了光栅扫描的概念；第12章讲述了视频中彩色编码的概念。本章结合光栅扫描和彩色编码的概念，形成了480i和576i的SD系统的基本技术参数。本章涉及的现代系统采用分量色彩，数字的 $Y' C_B C_R$ (BT.601)，或模拟的 $Y' P_B P_R$ 。第14章将对NTSC和PAL复合视频编码进行讲解。

13.1 扫描标准

两种扫描标准在世界不同地区的传统模拟电视广播中都有使用。480i29.97系统主要应用于北美和日本，今天大约占所有电视接收器的1/4。576i25系统主要用于欧洲、亚洲、澳大利亚和中美洲，大约占所有电视接收机的3/4。480i29.97（或525/59.94/2:1）俗称NTSC制式，576i25（或625/50/2:1）称为PAL制式。然而，NTSC和PAL只正确地应用到彩色编码中，而没有运用在扫描标准中。从扫描命名法看，两个系统之间的线数和场频是有差异的：在480i29.97视频中，场频严格是60/1.001Hz；在576i25系统中，场频严格是50Hz。

480i29.97和576i25数字视频的几个不同的标准如图13-1所示。

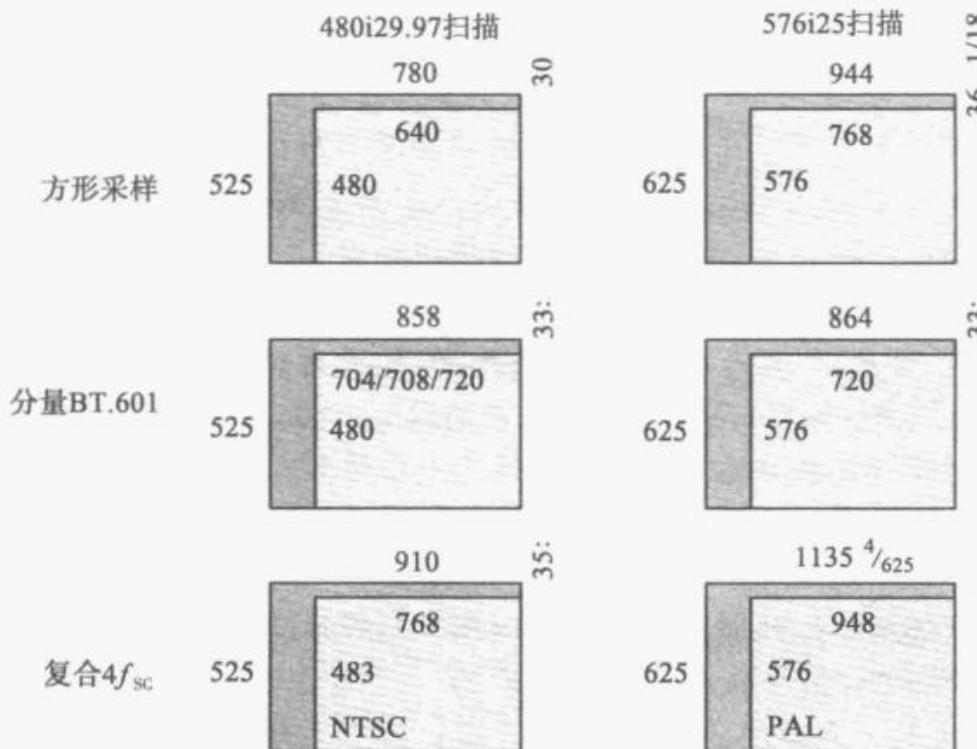


图13-1 4:3幅型比的SD数字视频光栅

480i的模拟广播系统通常采用约3.58MHz副载波的NTSC彩色编码；576i的模拟广播系统通常采用约为4.43MHz的彩色副载波的PAL彩色编码。使用一个从颜色中区分扫描的符号很重要，因为扫描和彩色编码的其他组合都在世界上重要的区域有所使用。巴西使用480i扫描和PAL彩色编码的PAL-M。阿根廷使用576i的扫描和与NTSC制式副载波几乎相同的3.58MHz的彩色副载波的PAL-N。在法国、俄罗斯等国家，使用SECAM

制式。符合这些难懂标准的设备已经不再生产：这些国家已经使用 480i 或 576i 标准的演播室设备，无论是在分量域或是 NTSC 制式的 480i 或 PAL 制式的 576i。这些演播室在播出前进行转码：颜色编码被改变，但不改变扫描，例如从 PAL 到 SECAM。

图 13-1 表明了每个标准的 S_{TL} 和 S_{AL} 值。 S_{AL} 值是一些复杂问题的结果，这将在 31.4 节进行讨论。欲了解有关 480i 系统的 483 有效行 (L_A) 的详细信息请参阅 31.3 节。

图 13-2 显示出了标准 480i29.97 和 576i25 数字视频的采样率，彩色编码通常与这些标准相关联。4 : 2 : 2 标准的 SD Y' C_B C_R 系统按照国际电联无线电通信部门（原 CCIR）的 BT. 601 标准进行了标准化。我把它称为 BT. 601。

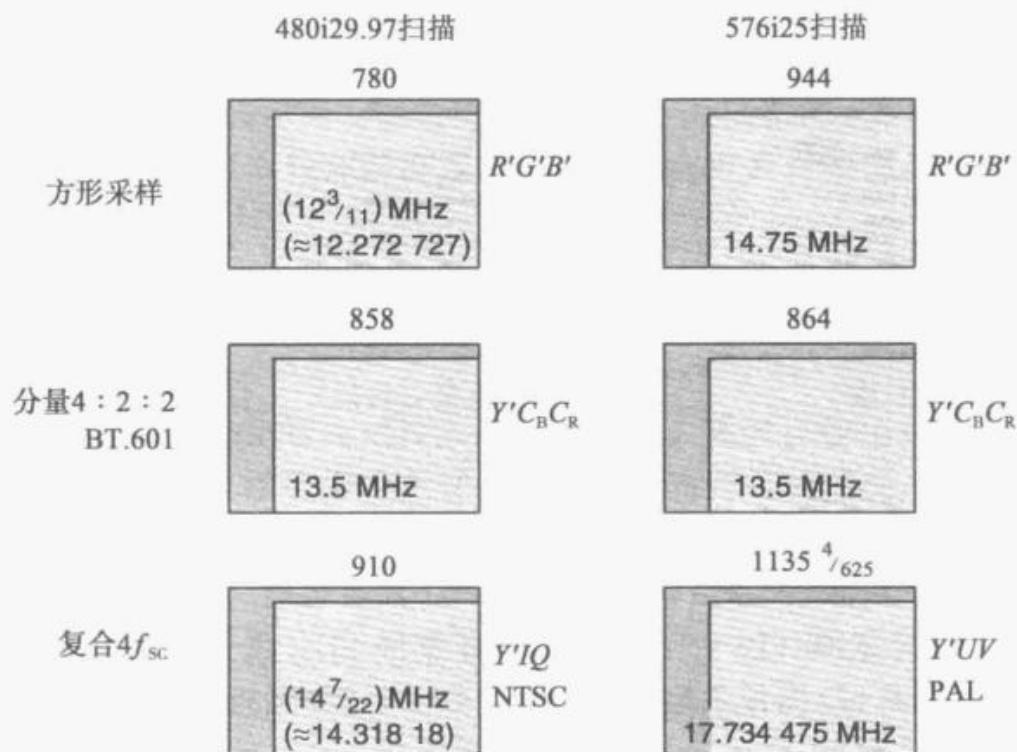


图 13-2 SD 采样率

除了一个例外，图 13-2 所示的所有采样系统有每个总线上的总样本数，这些系统是行锁定的。例外是复合 $4f_{sc}$ PAL 制式的采样，每总线样本为非整数 ($1135 \frac{4}{625}$) 的，这将会给系统设计师带来很大的麻烦。

480i 和 576i 有很多不同的技术参数，总结在表 13-1 中已给出。

表 13-1 480i 和 576i 之间的无偿差异

系 统	480i29.97	576i25
图像：同步比	10 : 4	7 : 3
建立，百分比	7.5	0
均衡、宽带脉冲的数量	6	5
行数 1 和 0 _v (定义在)	帧的第一个均衡脉冲	帧的第一个宽脉冲
底部图像行	第一场	第二场

480i 和 576i 不同的处理方法产生了图像数据不同的结构。这种差异在系统中的影响使得例如 MPEG 格式旨在同时适用于 480i 和 576i。图 13-3 和图 13-4 给出了字段顺序、交错命名和图像结构是如何联系的。图 13-5 显示了 MPEG-2 是如何识别每个字段的头部和尾部。在 480i 视频中，尾部段是帧的首部；在 576i 视频中，头部段是帧的尾部。图 13-3~图 13-5 描绘了图像阵列（即激活的样品），没有垂直消隐线，MPEG 对半线没有

规定。

图 13-3 所示图像中, 第一字段(以前称为奇数, 这里用 1 表示)以满图像行开始, 以包含图片尾部的左半线结束。第二字段(虚线处, 以前称为偶数), 大约 1/60s 以后被发送, 以包含图像的顶部右半行开始; 以满图像行结束。

图 13-4 所示图像中, 第一字段包括含图像顶部的右半线, 并以满图像行结束。第二个字段, 1/50s 后传送, 开始于一个满图像行, 以一个包含图像底部的左半行结束(在 576i 的术语中, 奇数和偶数很少使用, 并且最好避免)。

图 13-5 所示图像中, MPEG-2 隔行扫描根据是否包含该帧的顶部或底部来识别图像。顶部和底部的字段显示的是按照 MPEG-2 数据流编码的顺序。对于帧编码图像, 显示顺序是由第一字段的 1 位标志位决定的, 通常 576i 为 1, 480i 为 0。

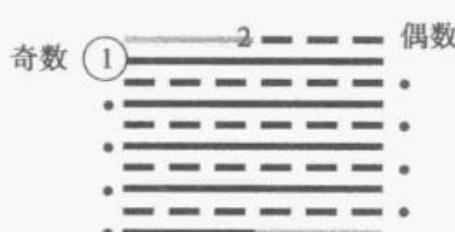


图 13-3 480i 隔行扫描

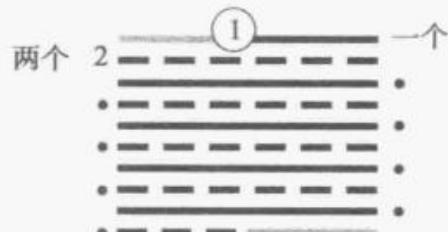


图 13-4 576i 隔行扫描



图 13-5 MPEG-2 隔行扫描

13.2 宽屏(16:9)SD

SD 中的编程用于以 4:3 幅型比显示。在 HD 发展之前(和过程中), 几种方案被设计出来, 以使 SD 适合宽屏(16:9)显示器, 即宽屏 SD。但这么说有误导, 因为没有增加像素数, 所谓的宽屏 SD 图像不能用比普通(4:3)SD 宽得多的图像角进行观察。(见第 7 章)。因此, 宽屏 SD 不能传送 HD 的主要原因是, 没有足够的宽阔的视角, 所以更准确的称呼应是宽幅型比 SD。再制定过渡期的各项方案现在已经过时, 这在《Composite NTSC and PAL: Legacy Video Systems》的“宽屏(16:9)SD”一节展开了讨论。

13.3 方形和非方形采样

计算机图形设备现在普遍采用方形采样, 即一个像素在横向和纵向上有等距间隔的采样格子。480i 和 576i 的方形采样图解如图 13-1 和图 13-2 所示。

虽然 ATSC 标准中的表 3 包括一个 640×480 像素的方形采样图像, 但没有演播室标准或实时接口标准用它来处理 SD 的方形采样。对于桌面视频应用, 我建议对于 480i 视频, 每总线严格用 780 个采样, 名义上有 $12\frac{3}{11}\text{MHz}$ 的采样率, 即 12.2727MHz 。为了适应演播室整个图像宽度, 需要 648 个采样; 通常 640 个采样用于 480 个图像行。对于 576i 视频方形采样, 我建议使用每总线 944 个采样, 即 14.75MHz 的采样率。

MPEG-1、MPEG-2、DVD 和 DV 都符合 BT.601 标准, 指定为非方形采样。480i 和 576i 的 BT.601 标准采样在图 13-1 和图 13-2 中已给出图解。

复合数字视频系统传统上以 4 倍彩色副载波频率($4f_{sc}$)进行采样, 于是就产生了非方形采样, 参数已在图 13-1 和图 13-2 的底部给出。正如我所说的, 复合 $4f_{sc}$ 系统已经过时。

在 480i 中, 方形采样、BT.601 和 $4f_{sc}$ 的采样速率成比例相关, 比例为 30:33:35。BT.601 的 480i 像素幅型比正是 $10/11$; $4f_{sc}$ 的 480i 像素幅型比正是 $6/7$ 。

在 576i 中，方形采样和 4：2：2 采样速率比是 59：54，所以 576i 的 BT. 601 的像素幅型比是 59：54。BT. 601 和 $4f_{SC}$ 采样率靠边沿比例相关联，对于数字硬件，这是相当令人费解的。

这种非方形采样大部分已经不再使用：大多数 HD 演播室标准要求方形采样，很难想象未来工作室标准会用非方形采样制定。

13.4 重采样

模拟视频可以简单地用适当的采样频率进行方形采样来数字化。然而，当进入方形采样桌面视频领域时，以 13.5MHz 标准数字视频采样率数字化的 SD 必须进行重采样，或内插，或按照 PC 说法为缩放。如果视频样本以 13.5MHz 传递到计算机图形系统，然后视为垂直和水平方向等距间隔处理，则图像的几何形状将被扭曲。BT. 601 480i 视频会出现横向拉长现象；BT. 601 576i 视频会出现压扁现象。在桌面视频中，往往在两个轴上都需要重采样。

比例 10/11 将 480i BT. 601 和方形采样联系起来：粗略亚采样可以通过简单地丢弃每个扫描线每第十一样本来完成！从 576i BT. 601 到方形采样的粗略采样可以通过复制每 54 个中的 5 个样本来完成（形式可能是 11-R-11-R-11-R-11-R-10-R，其中 R 表示一个重复样本）。然而，这样的样本丢弃和填充技术会引入混叠。我建议使用更复杂的内插器，这将在第 20 章进行解释。重采样可能沿垂直轴或水平（横向）轴进行，水平重采样较为容易，因为它按照光栅顺序处理像素，因此不需要任何行存储。

第14章

复合 NTSC 和 PAL 的介绍

在分量视频中，三个颜色分量是分开的。视频可以直接使用 $R'G'B'$ 分量，但三个信号的记录、处理或传输是代价比较大的。基于 $B'-Y'$ 和 $R'-Y'$ 的色差分量和亮度信号 (Y') 可以用来加强亚采样：亮度在全数据传输速率上得以保持，两个色差分量同时进行亚采样。即使在色度亚采样之后，视频仍然具有相当高的信息速率（数据传输速率，或“带宽”）。为进一步降低信息速率，复合 NTSC 和 PAL 彩色编码方案使用正交调制的方法，以使两个色差分量组合成一个单一的调制色度信号，然后采用频率交织的方法结合亮度和调制的色度成为具有大约 $1/3$ 数据传输速率的复合信号，或在模拟系统中，成为具有 $1/3$ 的 $R'G'B'$ 的带宽的信号。

复合编码是为了解决 3 个需求。首先要限制传输带宽。其次是为加强黑白接收机，该接收机在 1953 年规定了以最小的退化接收彩色广播。第三是新引进的彩色接收器应满足接收标准化的黑白广播的需要。在电视的初期复合编码是必要的，而且它已被证明对广播非常有效。NTSC 和 PAL 大概用于 10 亿的电子设备中。然而，分量数字视频已经超越复合技术，复合 NTSC 和 PAL 已是过时的技术。

相比分量数字视频，复合 NTSC 或 PAL 编码具有 3 个主要的缺点。首先，编码带来了一定程度的亮度和色度的相互干扰。一旦信号已经编码成复合形式，就会成为 NTSC 制式或 PAL 制式：信号中将引入交叉色度和交叉彩色误差。其次，不可能在复合域中直接执行许多处理操作，更何况重新定位或调整图像大小时需要解码、处理和重新编码。第三，如 JPEG 和 MPEG 等数字压缩技术不能直接应用到复合信号上，NTSC 和 PAL 编码的伪影对 MPEG 编码是有破坏性的。

单独传送颜色分量的带宽现在很容易实现，而且现在复合编码在演播室中已经过时。为了避免 NTSC 和 PAL 的伪影，便于图像处理和加强压缩，复合视频已经被分量数字视频取代，其中 3 个颜色分量 $R'G'B'$ 或 $Y'C_B C_R$ (数字系统)，或 $Y'P_B P_R$ (模拟系统) 是分开的。我希望你能避免接触复合 NTSC 和 PAL 制式，并跳过这一章节！

NTSC 和 PAL 正确地表示了颜色编码。然而它们经常被误用来表示扫描标准。PAL 编码已用于 576i 扫描（采用两个不同的子载波频率）和 480i 扫描（采用第三副载波频率）。单独的 PAL 是概念模糊的。

原则上，NTSC 或 PAL 彩色编码可以采用任何扫描标准。然而实际中，NTSC 和 PAL 仅采用 480i 和 576i 扫描标准，NTSC 和 PAL 编码参数在不同的扫描系统进行优化。本章介绍复合编码，详细介绍可参见《Composite NTSC and PAL: Legacy Video Systems》这本书。

14.1 NTSC 和 PAL 编码

NTSC 或 PAL 编码包括以下步骤。

- $R'G'B'$ 分量信号做矩阵变换并过滤，或 $Y'C_B C_R$ 或 $Y'P_B P_R$ 分量被缩放和过滤，以形成亮度信号 (Y') 和色差信号 (U 和 V ，在 NTSC 系统中为 I 和 Q)。
- U 和 V (或 I 和 Q) 色差信号被调制到一对密切相关的连续彩色副载波上，典型的

副载波是 480i29.97 标准下的 3.58MHz 或 576i25 标准下的 4.43MHz，以产生调制的色度信号 (C) (见图 14-1 左侧)。

- 亮度和调制色度相加形成复合的 NTSC 或 PAL 信号 (见图 14-1 右侧)。亮度和色度加和容易引入一定程度的相互干扰，称为交叉亮度和交叉色度。这些量可以通过频率交织最小化，下面即将进行说明。

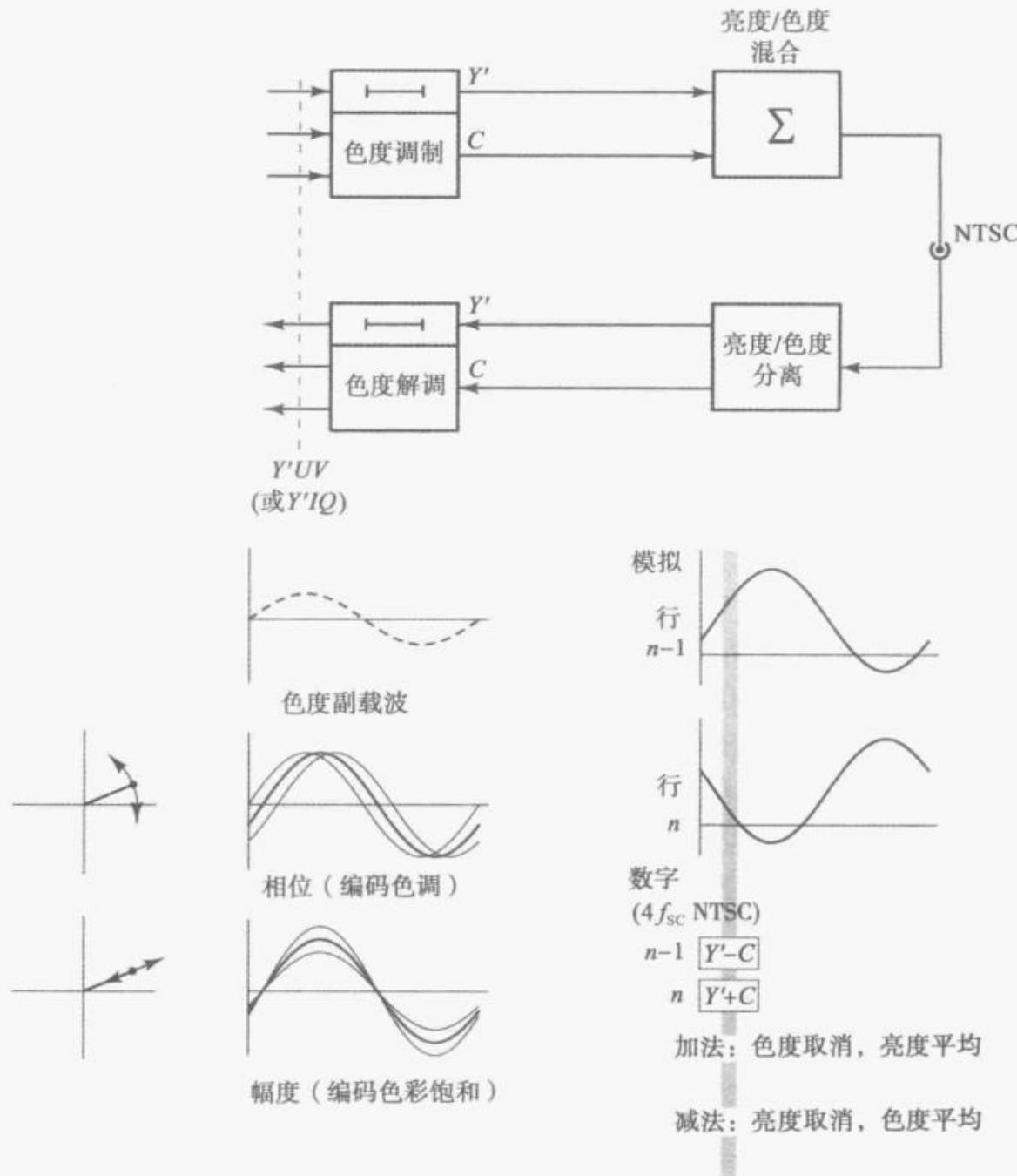


图 14-1 NTSC 色度调制和频率交织

图 14-1 中，连续地将亮度和一对色差分量编码为 NTSC 复合视频信号。首先，两个色差信号调制到彩色副载波上。如果两个色差是在极坐标系中，色调相角编码为副载波的相位，色彩饱和度被编码为副载波的幅度（未调制副载波的样本分帧后包括在复合信号中）。然后，调制色度与亮度加和。频率交织导致未调制的彩色副载波的逐个线相位反转，然后再导致调制的副载波的逐个线相位反转。相邻行的加和引起调制的色度中和，并且引起亮度平均。

14.2 NTSC 和 PAL 解码

NTSC 或 PAL 解码包括以下步骤。

- 亮度和调制色度被分离。粗分离可以使用一个陷波滤波器来实现。另外，利用频率交织实现大大改进的分离；在 NTSC 中，这样的分离器是一个梳状滤波器（在 S 视频接口，亮度和调制色度已经是分开的）。
- 解调色度产生 UV , IQ , $P_B P_R$, 或 $C_B C_R$ 基带色差分量。
- 如果需要求 $R'G'B'$ 分量，先对基带色差分量进行插值，然后亮度和色差分量进行反矩阵变换。

14.3 S 视频接口

S 视频涉及 NTSC 或 PAL 色度调制；然而，亮度和调制的色度通过接口穿过不同的路径而不是加和；S 视频接口绕过上述 NTSC 和 PAL 编码的第三步。交叉亮度和交叉色度的伪影要尽量避免。图 14-2 给出了编码器和解码器布置。S 视频接口广泛用在用户视频设备中。

14.4 频率交织

亮度和调制色度加和，会引入一定量的相互干扰。

通过频率交织使得干扰最小化，这是当彩色副载波频率和行频是一致时实现的，即当未调制的彩色副载波的相位锁定到仔细选择的合理行频的整数倍时——NTSC 的彩色副载波为平行频的整数倍，PAL 的彩色副载波为 $1/4$ 行频的整数倍。演播室中，通过从一个主时钟获取同步和彩色副载波来实现相干性。

在 NTSC 中，频率交织能够强化梳状滤波器的使用来分离亮度和色度：相邻行相加（形成垂直方向的平均亮度）和相减（形成垂直方向的平均色度），如图 14-1 的右下方所示。

在工业和消费类视频中，针对行频，子载波的随机性非常大，使频率交织的优势没能体现出来。大多数形式的模拟录像带记录引入时基误差，此误差未得到纠正，这也违背了频率交织的法则。

14.5 复合模拟 SD

复合模拟 NTSC 480i 和 576i PAL 制式广泛应用在消费电子设备（如电视接收机和录像机）中，并用于地面 VHF/ UHF 广播和有线电视已几十年了。详细信息请参见《Composite NTSC and PAL: Legacy Video Systems》这本书的第 12 章。

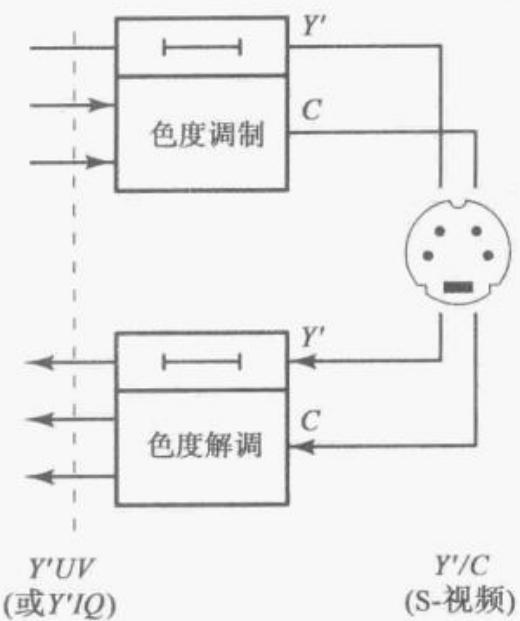


图 14-2 S 视频接口

第 15 章

HD 介绍

本章概述了高清晰度 (HD) 电视 1280×720 像素和 1920×1080 像素图像格式，并介绍了相关视频系统，比如 720p60 和 1080i30 的扫描参数。

如今的 HD 系统源于 NHK (日本广播公司) 的藤雄博士主持的一项研究。HD 具有相对于传统电视大约 2 倍的垂直和水平分辨率， $16:9$ 的画面幅型比，至少具有两个信道的 CD 质量的音频和 23.976Hz 或更高的帧频。按我的定义，HD 有 $3/4$ 百万像素或更高。NHK 设定 HD 采用隔行扫描；然而，逐行 HD 系统以后也出现了。

演播室 HD 采用 74.25MHz 的采样率，是 BT.601 标准下 SD 的 5.5 倍。HD 具有大约 6000 万像素每秒的像素率。除了少数令人讨厌的异常外， $R'G'B'$ 的编码参数类似于 SD 标准的参数，事实上，BT.709 标准下 HD 的一些参数适用于 SD 下的标准。关于扫描细节、采样率和 HD 接口电平的详情将第 41 章和第 42 章介绍，然而，HD 的 $Y'C_B C_R$ 颜色编码参数不同于 SD，详情将在第 30 章中介绍。

15.1 HD 扫描

20 世纪 80 年代和 90 年代一个关于 HD 应该采用隔行还是逐行扫描的重大讨论发生了。给定闪烁和数据传输速率，隔行扫描能够在静态空间分辨率上有所增加，如图 8-8 所示。广播电视传统上接受了伴随隔行扫描的运动伪影和空间混叠，以获得一些静态空间分辨率。在 HD 辩论中，计算机行业和创意电影界都坚决反对采用隔行扫描。最后，两个阵营在一定程度上妥协了，进而建立了 HD 标准以适应隔行和逐行扫描图像格式。在商业上可行的接收器必须解码这两种格式，虽然没有“合法”的要求规定要这样做。

图 15-1 给出 1 百万像素逐行扫描系统 (1280×720 像素，720p60) 和 2 百万像素隔行扫描系统 (1920×1080 像素，1080i30)，这两系统都符合标准。 1280×720 像素与 1920×1080 像素有非常简单的关系： 1280 是 1920 的 $2/3$ 和 720 是 1080 的 $2/3$ 。

除了 1 百万像素 (逐行扫描) 和 2 百万像素 (隔行扫描) 系统，一些 SD 扫描系统和一些额外的帧频被纳入最终 ATSC 标准，用于美国数字电视 (DTV)。表 15-1 总结了 ATSC A/53 标准的表 3 中发现的“18 种格式”。

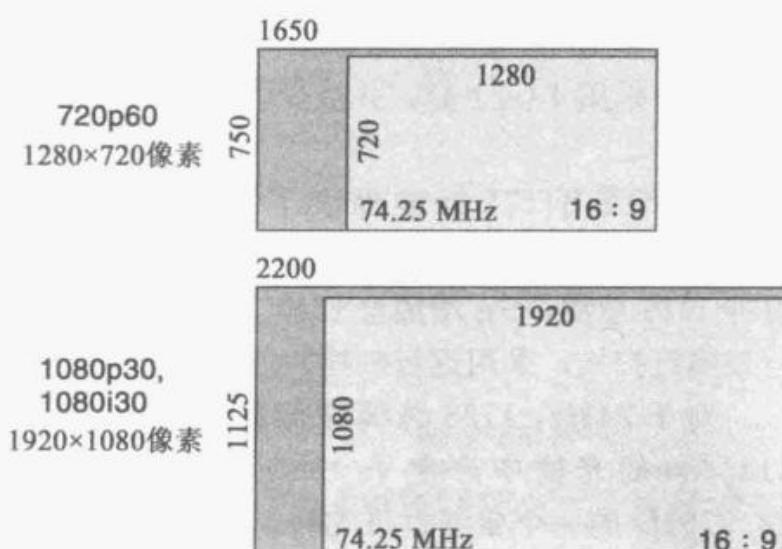


图 15-1 30 帧/s 和 60 帧/s 的 HD 光栅

表 15-1 ATSC A/53 标准的表 3

图像格式	逐行/隔行 ^① 帧频	图像幅型比
640×480	p 24, 30, 60 i 30	4:3 采样

(续)

图像格式	逐行/隔行 ^① 帧频	图像幅型比
704×480	p 24, 30, 60 i 30	4:3 方形
	p 24, 30, 60 i 30	16:9 非方形
1280×720	p 24, 30, 60	16:9
1920×1080	p 24, 30 i 30	16:9

① 帧频被比例 1000/1001 修正，也就是，帧频 23.976Hz, 29.97Hz, and 59.94Hz 是允许的。

ATSC A/53 标准的表 3 定义了针对美国数字电视所谓的 18 个格式，包括 12 个 SD 格式，作者发现 ATSC 标准的表 3 的布局有误，所以这里重新安排。ATSC 标准规定 704 S_{AL} 适用于几个 SD 格式，而不是 BT.601 的 720 S_{AL}（见 31.4 节）。ATSC A/53 标准不适用于 25Hz 和 50Hz 的帧频，但 A/63 标准适用。

1920×1080 系统被认为只采用隔行扫描 (1080i30)，但也可采用标准的 74.25MHz 的采样率进行 24Hz 和 30Hz (1080p24、1080p30) 的逐行扫描。24Hz 的适配器首次运用到数字电影。图 15-2 给出了 24Hz 和 25Hz 系统的栅格结构；表 15-2 总结了 HD 扫描参数。

在索尼旧式的 HDCAM 系统中，1920×1080 像素的图像下采样到 1440×1080 像素，色差信号在压缩之前按照 3:1:1 进行重采样。这只是一个内部表示；没有相应的未压缩的外部接口标准。目前索尼 HDCAM SR 格式以 4:2:2 直接表示 1920×1080 像素的图像数据（或在一些变型中采用 4:4:4），并减少了亚采样需要。

SMPTE ST274 标准提供了以 25Hz 的帧频传输 1920×1080 像素图像的方法：

保持 1125 总线，S_{TL} 增加至 2640。这产生了使用 1125/25 光栅 1080p25 格式。扫描可以是逐行或隔行扫描；采用逐行扫描的信号通常采用逐行分割帧 (PsF) 技术进行传输，已有介绍。

对于 24Hz，1125 总线被保留，S_{TL} 提高到 2750 以实现 24Hz 帧传输速率。这就在 1125/24 的光栅中产生了 1080p24 格式。这种系统主要用于数字电影（数字影院）。23.976Hz 的一个变型也可支持。

表 15-2 总结了 720p、1080i 和 1080p 系统的扫描参数。HD 演播室接口将在第 38 章介绍。

表 15-2 HD 扫描参数

系 统	扫 描	SMPTE 标准	S _{TL}	L _T	N _C (S _{AL})	N _R (L _A)
720p60	750/60/1:1	SMPTE ST 296	1650	750	1280	720
1080i30	1125/60/2:1	SMPTE ST 274	2200	1125	1920	1080
1080p60	1125/60/1:1	SMPTE ST 274	2200	1125	1920	1080
1080p30	1125/30/1:1	SMPTE ST 274	2200	1125	1920	1080
1080i25	1125/25/2:1	SMPTE ST 274	2640	1125	1920	1080
1080p25	1125/25/1:1	SMPTE ST 274	2640	1125	1920	1080
1080p24	1125/24/1:1	SMPTE ST 274	2750	1125	1920	1080

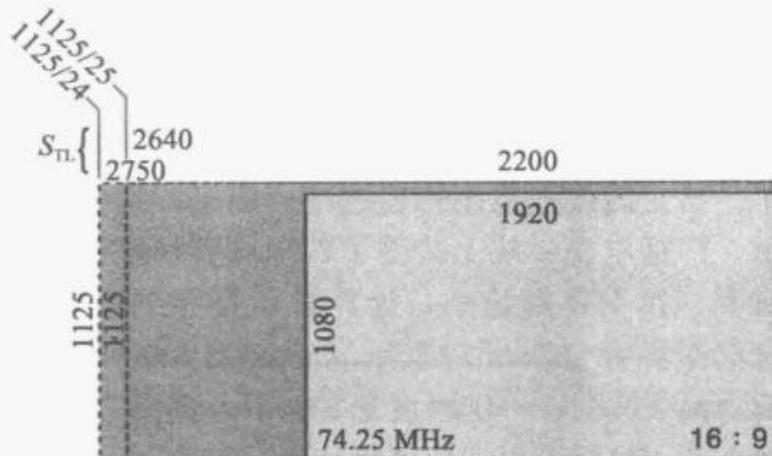


图 15-2 24Hz 和 25Hz 的 HD 光栅

15.2 BT. 709 HD 彩色编码

标准的 BT. 709 “编码” 传统观点涉及 OECF。第 2 章提出了，显示过程非常重要这一观点。完整节目内容的客观呈现需要标准化显示 EOCF。标准的摄像 OECF（如在 BT. 709）是工程用的，而对客观呈现没有影响。在实践中，所谓 BT. 709 节目内容要先建立 $R'G'B'$ 值，从而当这些 $R'G'B'$ 值通过 EOCF 参考例如 BT. 1886 标准转换成已知的观看条件（不幸的是，目前还没有标准化）下时，能够得到期望的图像外观。

BT. 709 定义了 $Y'C_B C_R$ 颜色编码。然而，BT. 709 标准下的亮度因子和由它们产生的 $C_B C_R$ 缩放因子不同于 SD。 $Y'C_B C_R$ 编码现在有两种形式：小图片（SD）编码和大图片（HD）编码。关于这个难题会在 28.10 节详细介绍。

第 16 章

视频压缩简介

直接存储或传输数字视频需要相当大的数据容量，SD 大概需要每秒约 20MB 的硬盘，HD 大约需要每秒 120MB 的硬盘。这里有一个关于存储容量和数据速率的拇指规则：2000 英尺的电影胶片卷轴可以携带一个 $133 \frac{1}{3}$ 分钟的电影；1 字节有 8 位，1 分钟有 60 秒， $\frac{60}{8} \times 133 \frac{1}{3}$ 是 1000。所以不论压缩与否，每秒 1Mb 相当于每部电影容量为 1GB。同样，每秒 1Gb 等于每部电影容量为 1TB。

经济的存储或传输方式需要压缩。本章介绍了 JPEG、M-JPEG、MPEG 和 H.264 压缩技术。

在前面的章节中，我们已经讨论了典型的在相当少的数字色彩分量（例如 3 个分量）的图像数据；每个分量有相当小数目的位（例如 8 位或 10 位）；通过非线性方式电光转换函数（EOCF）的感知编码和色度亚采样，可减少约 50% 的数据传输速率。在视频术语中，所有这些技术命名为无压缩视频，这也可能有不对的。压缩涉及变换技术，如离散余弦变换（DCT）和 JPEG2000 的情况下的离散小波变换（DWT）。

16.1 数据压缩

数据压缩用来降低用于存储或传送文字、数字、二进制图像、声音或其他数据所需的位数。高性能是通过利用数据的统计特性获得的。

位数降低是以一些计算压缩和解压缩为代价的。按照定义，压缩是无损压缩：解压缩按位（或字节）进行，数据呈现给压缩器压缩。

通用计算机应用程序的二进制数据通常有重复字节字符串的模式。大多数数据压缩技术，包括游程编码（RLE）和 Lempel-Ziv-Welch（LZW），利用重复字符串的优势实现压缩；性能高度依赖于被压缩的数据。

16.2 图像压缩

图像数据在相同颜色分量的样本中通常具有紧密的垂直、水平和空间相关性。当应用 RLE 和 LZW 算法计算存储在扫描线顺序的二层或伪彩色图像的数据时，一定程度上像素之间的横向关系可以被利用；这种技术通常会导致适度的压缩（或许 2 : 1）。

数据压缩算法可以设计为利用图像数据的统计信息，而不是任意的二进制数据；改进的压缩方法成为可能。例如，对于双层图像数据的符合 ITU-T 标准的传真，利用垂直和水平的相关性，可以达到比 RLE 或 LZW 更高的压缩比。按照定义，没有信道错误，数据压缩（甚至图像）是无损的：解压缩按位复原提交给压缩器的原始数据。

16.3 有损压缩

无损数据压缩可以进行优化，以实现连续色调（灰度或真彩）图像数据的适度压缩。但是，如果不要求精确重建，人类的感知特性可以被利用来实现明显更高的压缩比：图像

或声音数据可采用有损压缩，但引入的任何损失都不会察觉。有损压缩技术不适合用于二层或伪彩色图像；然而，它们对灰度或真彩图像和视频是非常有效的。

表 16-1 比较了用于 SD 和 HD 的 M-JPEG 和 MPEG-2 的压缩比。

表 16-1 近似压缩比率

格 式	没有压缩的数据速率 / (MB/s)	Motion-JPEG 压缩比	MPEG-2 压缩比	H. 264 压缩比
SD (480i30, 576i25)	20	15 : 1 (e.g., DV25)	45 : 1 (e.g., DVD)	90 : 1
HD (720p60, 1080i30)	120	20 : 1	75 : 1 (e.g., ATSC)	100 : 1 (e.g., 蓝光)

对于连续色调（灰度或真彩）图像，数据的压缩变换技术是非常有效的。在过去的几十年里离散余弦变换（DCT）已经被开发并不断优化；它是一种连续色调图像压缩选择方法。JPEG 指的是针对静态图像一种有损压缩方法。MPEG 指视频序列的有损压缩标准；MPEG-2 用在 DVD 和数字电视分配（例如，ATSC 和 DVB）。H. 264 指视频序列有损压缩标准。H. 264 对 HD 非常有效，它用在卫星、有线和电信（IPTV）系统，以及蓝光中。这些技术都将在随后的章节中描述。

16.4 JPEG

1992 年，JPEG 委员会通过基于 DCT 编码的标准，适用于压缩灰度或真彩色静止图像。这是万维网出现之前：该标准预计将用于彩色传真！JPEG 很快被采纳，并广泛用于桌面图形和数字摄影的静止图像中。M-JPEG 变体可用于运动序列；DV 格式采用类 M-JPEG 算法，第 45 章有介绍。

JPEG 压缩器通常转换 $R'G'B'$ 到 $Y'C_B C_R$ ，然后应用 $4:2:0$ 进行色度亚采样，以实现在变换编码步骤之前 $2:1$ 的压缩（在桌面图形，这 $2:1$ 的因子是包含在压缩比中的）。JPEG 具有规定直接压缩的 $R'G'B'$ 数据，无需二次采样。

16.5 运动 JPEG

JPEG 算法，虽然不是 ISO/IEC JPEG 标准，但已适应压缩动态视频。运动 JPEG 简单地压缩视频序列的每一字段或帧作为一个自包含的压缩图像，每个字段或帧是内编码的。由于图片被单独压缩，一个 M-JPEG 视频序列可以很容易地编辑；然而并没有利用时间相干性的优势。

视频数据几乎总是以 $Y'C_B C_R$ 亚采样形式呈现给 M-JPEG 压缩系统（视频中，因为色度亚采样而产生的 $2:1$ 因子不包括在压缩比中）。

M-JPEG 技术实现了从约 $2:1$ 至约 $20:1$ 的压缩比范围。SD 的 20MB/s 的数据传输速率可以压缩到约 20Mb/s ，适合于用户数字录像带记录（例如 DVC）。M-JPEG 的压缩比和磁带格式总结如表 16-2 所示。

表 16-2 M-JPEG 格式近似压缩率

压 缩 比	质量/应用	SD 磁带格式例子
2 : 1	“视觉无失真”演播室视频	数字广播级录像
3. 3 : 1	优秀质量演播室视频	DVC PRO50、D-9（数字 S）
6. 6 : 1	高质量演播室视频	D-7 (DVC PRO)、DVCA M、消费 DV

16.6 JPEG 2000

1995~2000 年, JPEG 委员会提出连续色调静止图像的压缩标准。这种努力最终成为 JPEG 2000 标准, 该标准是基于离散小波变换 (DWT) 技术而制定的。数字电影 DCI 标准使用 JPEG 2000 压缩标准。

JPEG 2000 的调整适用于运动序列, 其中每个 (逐行扫描) 帧单独编码, 不参考任何其他帧。虽然 “核心” 的 JPEG 2000 编码系统旨在成为权威和免许可证的, 但知识产权 (IPR) 的顾虑抑制了 JPEG 2000 的商业化。

16.7 夹层压缩

演播室中 DV 和它的衍生物是常见的。在过去的几年里, 适合采样和后期制作的一些基于软件的编解码器已经出现: Dirac/VC-2、DNxHD/VC-3 和苹果的 ProRes。夹层一词用于这样的编解码器, 符号化记录和用户分布之间的压缩率。

由 BBC 领导的一个开源开发项目开发出一名为 Dirac (为纪念诺贝尔奖物理学家) 视频压缩技术, 该技术主要用于有损夹层级的帧内小波压缩。Dirac 技术可以处理很多格式的视频, 但进行了将 10 位、4 : 2 : 2 1080p 视频压缩到数据传输速率约 50 Mb/s 和 165 Mb/s 之间的优化。SMPTE 已在 ST 2042 系列标准 (即 VC-2) 中对 Dirac 进行了标准化。

Dirac PRO 位流包含参数值, 通过使用配置文件和级别, 可以确定编码位流的复杂度。这些参数值使解码器能够很容易地确定它是否具有解码该位流的能力, 从而就定义了配置文件和级别。

Avid 采用很多的压缩系统并在后期制作上有广泛的应用。其中的一个是由 Avid 提出的 DNxHD, 已经由 SMPTE 标准化为 ST2042-1, 通俗地称为 VC-3。HD 的数据传输速率范围是从 60Mb/s 到 220Mb/s。压缩的数据通常被携带在 MXF 文件中。

苹果公司推行了一套专有的可变数据传输速率, 帧内基于 DCT 的编解码器叫 ProRes, 现在在采样和后期制作中使用相当广泛。ProRes 类似于运动 JPEG; 然而该方案是专有的, 细节没有公布。该计划旨在软件实现; 但是, 基于硬件的编码解码器是市售的。

一些帧内基于小波变换的编解码器已经被开发, 并广泛用于数字电影采样, 同时可以用于 HD。红色的红色代码是 JPEG 2000 的一种特定形态; GoPro 的 CineForm 同样采用小波压缩。这两种方案都在去马赛克之前分别对马赛克数据进行压缩。

图 16-1 所示的图像间编码利用了连续的视频图像之间的相似性。首先, 发送参考图像 (通常采用图像间压缩)。然后, 连续图像的像素差由编码器计算并发送。解码器通过累加像素差重构连续图片。假设该差值信息可以比原始图像信息更紧凑地进行编码, 那么该方案是有效的。

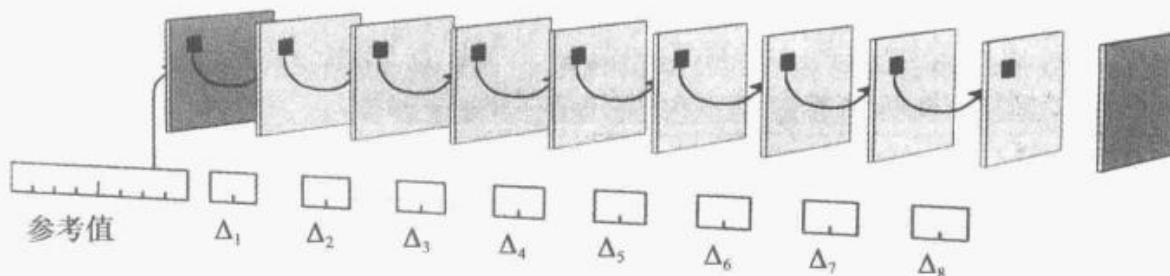


图 16-1 图像间编码

16.8 MPEG

除了场景变化，据统计，在视频序列中的连续图片非常相似。事实上，连续的图片大多是相似的，是有必要的：如果不是这样，人的视觉察觉不到图像序列！

变换编码效率通过利用视频的固有的时间冗余可提高 10 倍或更多。MPEG 标准是由运动图像专家组以在 ISO 和 IEC 为标准开发的。在 MPEG 中，帧内压缩根据后面图像预测提供一个初始的，自包含图像（参照图像）。然后编码器发送像素差，即预测误差，或者与基准的残差，如图 16-1 所示。该方法称为帧间编码（在隔行扫描的视频中，可以利用帧之间的差异，所以该方法更准确地描述为图像间编码）。

一旦解码器接受参考图像，就有了用于预测后续图像的基准。解码器接收到残差就能改善此估计。该方案有效的前提是，残差可以比原始图像信息更加紧凑地编码。

视频序列的运动将导致场景中画面相对于所述图像阵列的移位。快速移动的图像很容易在一帧时间中产生 20 个像素的移位。运动的存在，像素在某个固定位置与在连续图像中灰度值可能有完全不同的值。运动易于引起预测误差信息，并且与像素点大小正相关，这样就显示不出码间编码的优势。

图 16-2 中，GoP 有 9 张图片，编号为 0~8。其中，9 是下一个 GoP 的第一图像。I 图像 0 是从表示为绿色块的编码数据进行解码的。这里，帧内数 (n) 为 9。

然而，即使存在移动，图像也会保留自己的空间结构。MPEG 采用运动补偿预测 (MCP) 克服了图像之间的运动的问题。编码器配备有运动估算 (ME) 电路，用于计算运动向量。编码器然后由估算的运动（一种称为运动补偿的内插处理）重置参照图像的像素值，然后由运动补偿的参考计算残差。编码器使用类 JPEG 技术压缩残差，然后发送运动向量和压缩的残差。

基于接收到的运动向量，解码器模拟编码器的运动补偿内插，得到的预测值比未重置的参考图像效果更好。所接收到的残差应用（通过简单的加法）到重建编码器的近似图像。

该 MPEG 系列标准规定了标准的位流的属性和由解码器执行的算法。产生兼容（“合法”）位流的任何编码器认为是属于 MPEG 标准的，即使它产生的是质量差的图像。

16.9 图像编码类型 (I、P、B)

在 MPEG 中，将视频序列分割成图像组 (GoP)。在每个 GoP 的第一图像是使用类 JPEG 算法编码，独立于其他图片。这是帧内或 I 图像。一旦重建，一个 I 图像作为基准图像可用于预测邻接（非组内）图片。图 16-2 给出的 GoP 例子包括九张图片。

P 图像包含从临近的锚帧预测的参数。一旦 P 图像重建，就会被显示；此外，它成为一个新的锚。I 图像和 P 图像形成两层分层结构。图 16-3 给出了一个 I 图像和两个相关的 P 图像。

MPEG 提供了一个可选的第三层级，由此 B 图像可插在锚画面之间。B 图像的元素通常是这样预测的，从过去的参考图像的运动补偿单元和未来的参考图像的运动补偿单元进行平均得到（在编码器判决时，B 图像的参数可以是从前面的参考进行单向前向插值，或

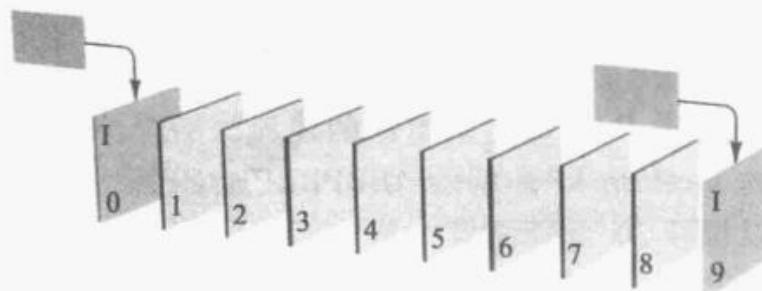


图 16-2 MPEG 图像组 (GoP)

从后面的参考进行单向向后预测)。每个 B 图像进行重构, 显示, 然后弃去: 未解码的 B 图像形成任何预测的基础。以只用 I 图像和 P 图像进行编码相当的压缩率, 利用 B 图像传送大幅增益, 但会引入编码延迟。

图 16-4 示出了两个 B 图像。

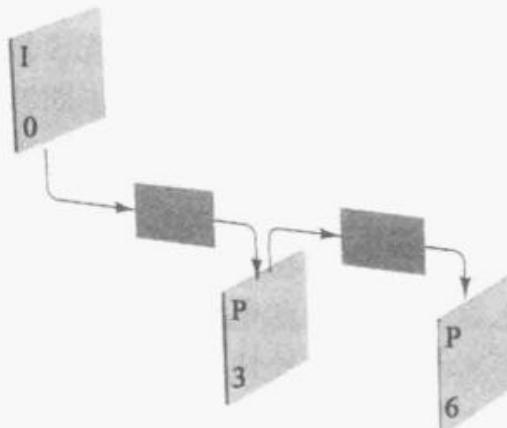


图 16-3 MPEG P 图像

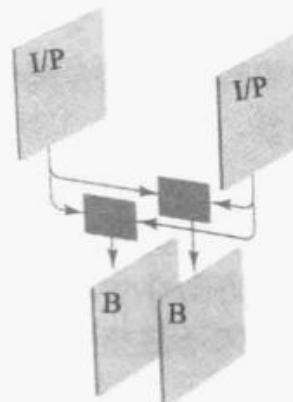


图 16-4 MPEG B 图像

图 16-5 总结了三级 MPEG 图像层次结构, 这个例子具有结构 IBBPBBPBB。

一个简单的编码器通常产生具有固定顺序的 I、P 和 B 类型图像的位流。一个典型的 GoP 结构表示为 IBBPBBPBBPBBPBB。以每秒 30 个画面算, 每秒就有两个这样的 GoP。周期性的 GoP 结构, 可以用一对整数 N 和 M 来说明; N 是一个 I 图像 (包含的) 到下一个 (不包含的) 的图像数目, 而 M 是从锚定图片到下一个锚定图片的图像数目。如果 $M=1$, 不存在 B 图像。图 16-5 展示的是一个常规的 GoP 结构, 其中, 时间间隔 $N=9$ 的 I 图像, 且 $N=3$ 的锚定图像间隔。 $M=3$ 分量表示锚画面之间有两个 B 图像。很少超过两张 B 图像处于参考图像之间。

GoP 中编码的 B 图像依赖于 P 图像和 I 图像; 编码的 P 图像依赖于更早的 P 图像和 I 图像。由于这些相互依赖关系, 除了在 GoP 边界, MPEG 序列不能被编辑, 除非该序列被解码, 编辑, 并随后再编码。MPEG 非常适合分配, 但由于其不能在任意点进行编辑而不失真, 所以 MPEG 通常不适合用于生产。在仅用于 MPEG-2 的 I 帧的 MPEG-2 专业领域中, 每一个 GoP 是一个单一的 I 帧。这在概念上等同于动态 JPEG, 但对于一个国际标准是大有好处的 (MPEG-2 的另一种变型中, 简单地说, 就是没有 B 图像)。

我已经介绍了 MPEG, 貌似每一个 P 图像中的所有元素和每一个 B 画面中的所有元素都以类似的方式编码。即使是通过过去参考图像很好地预测得到的图片也存在无法有效地预测的几个区域。在 MPEG 中, 图像被拼接成 16×16 亮度样本块, 并且编码器是考虑到代码的任何特定的宏块帧内模式的选择, 即独立于任何预测。对于一个宏模块应该跳过的小型代码信号, 运动补偿预测可以使用并无需修改。在 B 图像中, 编码器可以使用前向

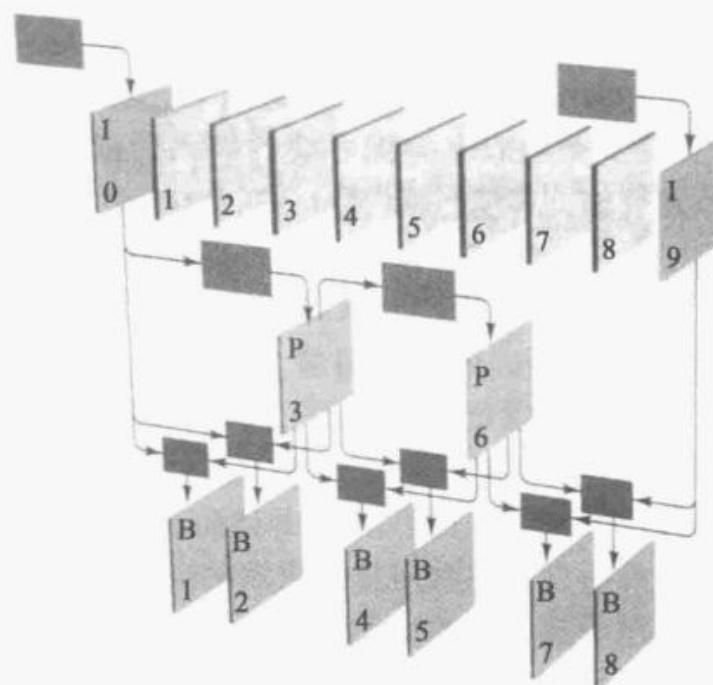


图 16-5 三级 MPEG 图像层次结构

预测，后向预测，或双向预测的宏块，以宏块基础决定代码。从形式上看，I 图像只包含 I 宏块；一个 P 图像具有至少一个 P 宏块，一个 B 图像具有至少一个 B 宏块。

16.10 重排

在无 B 图像的序列中，I 图像和 P 图像进行编码，然后将其存储或以明显顺序发送。然而，当使用 B 图像时，解码器通常需要访问过去的锚定图片和将来锚定图片重建一个 B 图像。

图 16-6 所示的是编码器压缩序列（其中锚图片 I_0 , P_3 和 P_6 以粗体显示）。编码 B_1 和 B_2 的图像可能从 P_3 进行后向预测得到，因此编码器必须缓冲未压缩的 B_1 和 B_2 图片直至 P_3 编码：只有当 P_3 编码完成， B_1 才能开始。使用 B 图像导致编码延迟（见图 16-7，序列进行编码没有 B 图片，编码信息 P_1 传输就不会受到这两个画面延迟的影响）。编码延迟（时延）可以使存在 B 图像的 MPEG 不适合实时双向应用，如电话会议。

图 16-6 所示的编码 9 图像 GoP，若按显示顺序发送，则该译码器就必须保存编码 B_1 和 B_2 的数据在缓冲区，同时接收并解码 P_3 ；只有当 P_3 的解码完成， B_1 才能够开始解码。无论如何，不管怎样最小化缓冲存储器，编码器都必须缓冲 B_1 和 B_2 ，MPEG-2 指定编码 B 图像信息在编码的参考图像之后传输。图 16-8 表示图像数据的重新排序，以便传输。我已将 I_9 标在括号中，因为它属于下一个 GoP (GoP 头在它前面)。在这里， B_7 和 B_8 紧跟 GoP 头。

$I_0 B_1 B_2 P_3 B_4 B_5 P_6 B_7 B_8$

图 16-6 GoP 示例

$I_0 P_1 P_2 P_3 P_4 P_5 P_6 P_7 P_8$

图 16-7 无 B 图像 9 图像 GoP 示例

$I_0 P_3 B_1 B_2 P_6 B_4 B_5 (I_9) B_7 B_8$

图 16-8 为传输重排序的 GoP

16.11 MPEG-1

对原 MPEG 标准改进导致了现在所谓的 MPEG-1，这是在多媒体中应用的。MPEG-1 是对逐行 352×240 像素图像以 30 帧/s (240p30) 的编码进行优化。MPEG-1 对隔行扫描图片没有规定。当 480i29.97 或 576i25 视频用 MPEG-1 以典型的数据速率进行编码时，每个帧的第一场被编码，好像它是渐进的，并且所述第二场被舍弃。在大约 1.5Mb/s 的预期的数据速率下，MPEG-1 提供 VHS 品质的图像。

对于广播，MPEG-1 已被 MPEG-2 取代。MPEG-2 解码器必须解码 MPEG-1 的参数限制位流 (CPB) 序列，这在表 47-1 中讨论，所以我不会进一步讨论 MPEG-1。

16.12 MPEG-2

在 MPEG-2 的开发中，将 MPEG-1 延伸到隔行扫描，延伸到更大的图片，并且采用高于 1.5Mb/s 数据速率。MPEG-2 进行标准化是基于 ISO/IEC 的一系列文件；MPEG-2 广泛应用到数字电视 (DTV)，包括 SD 和 HD (例如 ATSC)，并且采用针对 DVD 的视频压缩方案。

MPEG-2 可以适用于逐行扫描视频和隔行扫描视频。一个视频帧可以直接被编码为帧结构的图像。另外，一个视频帧（通常来自一个隔行的源）可以被编码为一对场构造的图像——顶部场图像和底部场图像。这两场是时间偏移一半帧的时间，并且都是用于隔行显示的。场图像总是具有相反的奇偶性成对的出现（顶部/底部或底部/顶部）。在一个场对中，两个图像具有相同的图像编码类型 (I、P 或 B)，不同的是 I 图像之后可以是一个 P 图像（在这种情况下，这个场对有效地作为一个 I 帧）。

以下介绍的 MPEG-2 变型，MPEG IMX，主要用于演播室。MPEG-2 的变型，HDV，主要用于大众用户。MPEG-2 视频压缩从第 47 章开始详述。

16.13 其他 MPEG 标准

当 MPEG-2 的工作正在进行时，一个解决高清问题的 MPEG-3 正在推出。MPEG-3 委员会很早就得出结论，MPEG-2 在高数据率方面能满足高清的需要；因此，MPEG-3 被放弃了。下面我将讨论 MPEG-4。MPEG 数字大于 4 是善变无常的。

MPEG-7，冠名为多媒体内容描述接口，规范着各种类型的多媒体信息（元数据）的描述。在我看来，MPEG-7 是与处理工作室或分发高质量视频信号并不相关的。

根据 ISO，MPEG-21 “定义了一个开放的框架，用于多媒体传输和使用，同时以内容创建者和内容消费为重点。该设想的 MPEG-21 是定义一个多媒体框架，使在广泛的网络和使用不同的社区设备的透明和增强运用多媒体资源。”在我看来，MPEG-21 与处理工作室或分发质量视频是不相关的。

16.14 MPEG IMX

索尼用于 SD 的 480i 和 576i 原装数字磁带，采用的是专有的类似运动 JPEG 压缩。第一个产品是磁带录像机用 Betacam SX 表示，具有大约 18MB/s 的数据传输速率。采用 I 帧的 MPEG-2422P@ML 的后续产品，记为 MPEG IMX，具有 30Mb/s、40Mb/s 或 50Mb/s 的数据传输速率 (IMX30, IMX40, IMX50)。之后 MPEG IMX 磁带录像机逐渐商业化。今天在一个 MXF 文件中放置或包裹 IMX 压缩视频是很常见的。

XDCAM 系列是索尼指定的使用各种压缩系统和多种物理介质的产品。MPEG IMX 压缩是 XDCAM 提供的压缩系统中的一个。记录在光盘介质是可能实现的。

16.15 MPEG-4

MPEG-4 最初目标是以非常低的数据传输速率进行视频编码。这导致了视频压缩系统标准化为 MPEG-4 第 2 部分视觉；它不同于 MPEG-2 和 H.264。ISO/IEC 共同开发 ITU-T H.264 标准作为 MPEG-4 部分 10，因此单独 MPEG-4 一词是不明确的。

MPEG-4 第 2 部分界定高级简单类 (ASP)，由 DivX 和 Xvid 实现。ASP 对专业质量的视频没有用。即使是在 ASP 预期的应用领域，低数据传输速率的视频（在一段时间称为 H.264）已被证明具有更好的性能。因此，ASP 已经失宠。

MPEG-4 第 2 部分还定义了一个名为简单工作室类 (SStP) 的配置文件。此文件以非常高的数据传输速率应用在索尼的 HDCAM SR 中 (MPEG-4 被设想的数据传输速率频谱的另一端)。无论是在磁带和文件中，HDCAM SR 广泛应用于 HD。除了 HDCAM SR，MPEG-4 简单工作室配置文件用途非常有限。

MPEG-4 系列标准第 12 部分规定了一个基于时间的媒体文件通用容器结构的 ISO 基本媒体文件格式。该格式采用的是桌面视频（最常见的在 MP4 文件中的 MPEG-4 第 2 部分/ASP 视频），但是很少用于专业视频分发。

16.16 H.264

扩展 MPEG-2 编码的工作是由联合视频组 (JVT) 负责的。在开发过程中，关于这方面的工作是朝着高级视频编码 (AVC) 方向努力的。由此产生的标准是由 ITU-T (称为 H.264) 和 ISO/IEC (尽管它与 MPEG-4 的其余部分有很少的关联，人们仍称它为

MPEG-4 第 10 部分) 共同颁布的。H.264 编码大约比 MPEG-2 高效 1.5 或 2 倍, 也就是说, H.264 通常允许编码在稍微超过类似图像质量水平的 MPEG-2 的一半数据传输速率。蓝光标准规定了用户播放器的 H.264 解码的内容。详细信息请看第 48 章。

16.17 帧内 AVC

一个只有 I 帧的 H.264 标准已经引用到专业视频采样和制作, 称为帧内 AVC。10 位视频以 720p, 1080i 或 1080p24 格式压缩为 50Mb/s 或 100Mb/s 的数据传输速率。

在帧内 AVC 50, 1920×1080 像素格式中, 图像在压缩前亚采样为 1440×1080 像素, 4:2:0 格式; H.264 Hi10P 帧文件在第 4 级时被使用。1280×720 像素的图像亚采样为 960×720 像素, 4:2:0 格式; H.264 的帧文件在 3.2 级时被使用。

帧内 AVC 100 格式在本地对 1280×720 像素和 1920×1080 像素格式进行编码, 使用 4:2:2 色度亚采样, 在 4.1 级进行 H.264 Hi422P 文件的编码。

16.18 WM9、WM10、VC-1 编解码

微软开发了一系列的视频编解码器, 而效果最好的是 Windows 媒体视频 9 和 10 (WMV9 和 WMV10)。这些编解码器在概念上类似于 H.264——但是在细节上不同于 H.264。微软将 WMV9 编解码器提交给 SMPTE, 对它作了略微的改动, 标准化为 VC-1 (这是与 Windows 媒体视频 9 (WMV9 或 WM9) 演播室配置符合的)。

微软针对 WM9 (及其下一代 WM10) 开发了一个专门的编码器和解码器。VC-1 规范用来产生一个解码器的“洁净室”, 其参考源代码是开源的。

在用户播放器中蓝光标准规定了 VC-1 的解码能力, 并且数以百计的蓝光光盘的也是以 VC-1 标准完成的。如果你正在开发蓝光播放器, 那么你就需要实现 VC-1 解码。然而, 微软却停止了这一系统的开发。现在 VC-1 处于奄奄一息的状态, 后面对 VC-1 将不再作进一步的讨论。

16.19 CE 采样压缩

传统的消费类电子设备是基于硬件的。当今两个视频压缩格式主宰了消费电子 (CE): 基于 MPEG-2 的 HDV 格式和基于 H.264 的 AVCHD 格式。然而, 从 IP 网络上的视频显示到个人计算机上, 现在已经有了巨大的经济价值。稍后我将讨论 IP 传输和基于软件的解码优化压缩技术。

16.20 HDV

HDV——高清视频, 指的是适合于消费电子设备的压缩系统。采用长 GoP 帧间 MPEG-2 压缩时, HDV 与 Y'CBCR 8 位 4:2:0 视频有不同的帧频。

720p 的 HDV (有时称为 HDV1) 提供 25, 30, 50Hz, 或 60Hz 帧频, 具有 1280×720 像素的图像结构和 4:2:0 色度亚采样, 按照 MP @ HL 压缩, 并以约 19Mb/s 的数据传输速率记录一个 MPEG-2 传输流 (TS) 和一个短的 (6 帧) 的 GoP。

25Hz 或 29.97Hz 的 1080i HDV (有时称为 HDV2) 随着 4:2:0 色度亚采样下采样到 1440×1080 像素, 根据 MP@H14 压缩, 并以大约 25Mb/s 的数据传输速率记录分组基本流 (PES) 与长 (15 帧) GoP。亮度采样率是 55.6875MHz, 即工作室标准 74.25MHz 的四分之三。

HDV 适应 MPEG-1 层 II 立体声音频。用户摄像机可以使用 MiniDV 磁带介质, 硬盘

介质和闪存介质。

16.21 AVCHD

H.264 中专门面向用户使用的叫 AVCHD。AVCHD 格式压缩的 720p, 1080 i 和 1080P24 视频, 采用长 GoP H.264 编码, 数据传输速率大概在 6Mb/s 和 18Mb/s 之间。杜比数字音频编码也被采用。在 12cm DVD-R 介质, SDHC 闪存卡, 或硬盘驱动器上记录 AVCHD, 已经达到了半专业水平。

16.22 IP 传输到用户的压缩

个人计算机、笔记本电脑、手持设备和平板电脑的激增, 大多数具有 WiFi 连接, 并且都能够进行基于软件的视频解压缩, 导致基于软件的编解码器的实现, 并且性能已经超越 MPEG-2。

苹果公司已经投资并且已经广泛应用 H.264 编解码器。H.264 在本章前面已有概述, 并将详细说明。

谷歌采用 WebM, 下面将概述。

这个应用领域有少数专有编解码器, 有 RealPlayer 和 Adobe Flash 播放器 (最近的 Flash 播放器包括 H.264 解码)。

16.23 VP8 (WebM) 编解码

2010 年, 谷歌收购了一家经营了 10 年或更长时间名为 On2 的公司, 开发了一系列用于视频分配的专门的基于软件的编解码器。谷歌以开源的方式发布了 VP8 编解码器, 并将 WebM 系统推广到视频网络分发 (基于 IP) 中。WebM 压缩用 VP8 编码的视频和用 Vorbis 编码的音频, 都是打包到 Matroska 文件中。基于 MPEG-2 或 H.264 的知识产权 (专利) 许可证, 谷歌的目的是避免支付专利费, 并避免其客户支付特许权使用费。

16.24 狄拉克 (基本)

基本狄拉克是由 BBC 和其合作伙伴开发的长 GoP 运动补偿的帧间小波编解码器。它已由 BBC 广泛应用在 2~4Mb/s 之间的 SD 分发和 15~18Mb/s 之间 HD 分发中。

第17章

流和文件

文件是有明确的开头和结尾，以存储信息为特征的有序位序列。流则是以在单向信道中、无界的持续时间上实时数据传输为特征的结构——也就是说，没有上行信道用于流的控制、确认或者请求重发。表 17-1 给出了文件和流的特征概括。

表 17-1 文件和流的比较

文 件	流
…含有预先定义好的开头和结尾	…开头和结尾是模糊的
…通常包含存储媒介	…通常包含外部数据关联
…允许随机访问数据	…包含序列数据访问，通常是从中间流开始访问
…在高层有固定结构；数据可以任意安排	…在低层有固定结构；安排数据以用来减小缓冲
…不需要内嵌分隔符	…包含内嵌分隔符，使得在飞速写入时关键元素能被识别
…传输通常通过一般的网络进行	…传输通常通过数据互连进行
…传输通常是不同步的	…传输通常是时间同步的
…传输通常有可变的数据传输速率（VBR）	…传输通常有固定的数据传输速率（CBR）
…传输数据的完整性有保证，但是传输速率没有（只能尽力而为）	…传输速率有保证，但是传输数据完整性却没有（误码时有发生）
…传输通常有上行通信；传输通常会被应答	…传输通常没有上行通信；传输通常也不需应答

流是“直播”的，它适合实时连接。文件在本质上并不是“直播”的。文件的调用或交换可能会慢于实时、相当于实时、或者快于实时应用。流的一部分可以被存储为文件，而文件也能以流化方式通过接口。但是，总的来说，流是为实时通过接口而设计的结构，文件是在存储介质上为非实时应用设计的结构。一般来说，视频接口（还有磁带录像机）是以流为特征的，视频存储是以文件为特征的。

17.1 历史概述

视频信号一直以来都是设计成流的形式，磁带录像机记录了连续的流（传统上会忽略垂直消影间隔，在数字录像机中，也会忽略水平消影间隔）。

流接口在飞速传输元素（类似于以前的模拟视频同步）时允许同步：接收端可以在任何时候连接上流，然后在零点几秒后开始操作。流的格式被设计成具有局域性，借此关键元素可以同时显示。通常情况下，视频和相关音频，在流中处在相近的位置，这样可以限制对缓冲存储器性能的要求，限制显示时对关键元素的访问延时。

以前，未经压缩的数字视频在演播室中会以流的形式实时通过数字串行接口（SDI）和高清数字串行接口（HD-SDI）。SDI 和 HD-SDI 的定时设计为，模拟视频可以通过剥离流的同步和辅助元素、执行数/模转换和插入模拟同步而简单地获得，几乎不需要缓冲存储器。

由于数字录像和压缩数字视频的回放成为了现实，SDI 中的串行数字传输接口

(SDTI) 被专门设计用来“包裹”压缩视频，音频也通过 SDI 传输，各种压缩方案例如 MPEG IMX 和 DV100 都适用。数字视频能通过 IEEE 1394 接口以流模式传输。

随着计算机和网络技术的发展，压缩视频的存储逐渐可行，然后是非压缩视频以文件形式进行存储。通过 TCP/IP 以太网进行这类文件交换变得越发普遍，起初以 100Mb/s 的数据传输速率进行传输，接着数据传输速率达到 1Gb/s，以至于到 10Gb/s。

随着信息网络技术的发展，为了通过特殊接口传输而打包压缩视频的方案例如 SDTI 已经应用于通用网络中。如格式为 MPEG IMX 和 DV100 的压缩视频是以原始位流形式存储在文件中的。素材交换格式 (MXF) 是作为一种机制出现的，在更多架构中用于存储视频和音频的关键内容。相应的标准则用于以 MXF 文件形式存储压缩视频。MXF 文件不需要包含关键内容：它可以让关键内容存储到分离的文件中。在 MXF 文件中一些以 A/V 存储的格式就本质而言是“流型”的，视频和音频的关键数据保存于 MXF 文件相近的位置，以便以最小重组结构播出。其他具有更高级结构的格式更适合于后期制作（例如，将视频和音频存储在由主 MXF 文件指定的分离文件中）。

如今，基于文件的工作流程广泛应用于视频制作和后期制作，然而，基于流的技术仍然在专业制作内容的分发上占主导地位。像 Youtube、Hulu、Netflix 等公司提供的服务，将视频通过本书所说的“大绒毛互联网”分享给用户——然而，这种系统的服务和质量水平会比相关的电视广播的水平要低：画质不是高清水准；视频出现停顿，音频不同步。互联网协议电视 (internet-protocol television, IPTV) 是指用基于 TCP/IP 的商业网络来达到广播的服务和质量水准的。

苹果的 Itunes 模式不同：它是在文件传输后播放的。

在专业的视频分发中，基于文件的制作和后期制作模块，以及基于流的分发模块都会出现在播放服务器中，播放服务器含有使用关联文件系统的磁盘存储机制。在制作时，服务器使用互联网进行异步存取。在发送时，流存取机制根据由演播室同步和时间编码驱动的时间轴来读取文件，并进行相应的回放。可能会用到对压缩视频的实时解压，专用的流接口会将内容载入到分发网络中。

17.2 物理层

串行数字视频 (Serial Digital Video, SDI) 接口和 HD-SDI 接口是基于 10 位的码字，经串行化和“加扰”后以位流的形式单向地在单总线上传送。加扰技术允许有效负载的数据传输速率等于总线上的数据传输速率；然而，信号同步需要在视频中禁止某些数据值的出现。

商业计算机接口都是基于 8 位字节的。以前，在单导线上数据是串行的（如以太网）；但是很多面向计算机的接口（如 PCIe 接口和雷电接口）使数据在多条线路上实现串行化。在一些物理接口上，数据通常会用例如 8 位/10 位方案进行映射——使得所有 8 位字节值可以通过该接口传送，同时允许接收机对于任意数据能够恢复时钟和数据帧。在 8 位/10 位的映射情况下，信道的数据传输速率如编码率要比有效载荷率略高，为 1.25 倍。其他接口（如 DVI 和 HDMI）使用专用的时钟线路来建立定时；任何数据可以串行化和传输而不受数值的限制。一些接口通过每个导体（或每对导体）将数据传输方向反向，其他的接口每个传输方向均有专用的导线（或导线对）。

17.3 流接口

SDI 接口是为 4 : 2 : 2 的未压缩标清视频而设计的，它拥有 270Mb/s 的数据传输速率

(SDTI 采用 SDI 以用于压缩视频，只不过 SDTI 现在差不多被废弃了)。HD-SDI 接口是为 4:2:2 的未压缩高清视频而设计的，它拥有约 1.5 Gb/s 的数据传输速率能力。近年来，数据传输速率为 3 Gb/s 的经改进的 HD-SDI 已经规范化和商用化了。对于 SDI 和 HD-SDI 的详细介绍请参见第 38 章。

DVI、HDMI 和显示端口 (displayport) 是专门用来为连接计算机图形子系统显示而设计的数字接口，其电缆长度可以达到 3m。除了一条速率非常低的反向通道，即显示数据通道 (display data channel , DDC)，将显示特征的上行流传送给图形子系统之外，DVI 和 HDMI 都是单向的。

DVI 为交互式数字视频系统，HDMI 为高清晰度多媒体接口，DisplayPort 为显示接口。

雷电接口是由英特尔公司设计的双向接口 (显然是与苹果公司合作的)。这一方案试图将计算机与外围硬件连接起来：其单向数据传输速率可以达到 10 Gb/s，驱动电缆可长达 3m，雷电接口的链接可以是菊花链式的。显示端口的提出要早于雷电接口；然而，显示端口 (包括迷你显示端口接口) 是雷电接口的特例，除了用极低速率的反向信道作为显示数据通道之外，数据流都是单向的。雷电接口可以传输两路未压缩的 1080p、60 帧的 R'G'B' 集。数据流以包的形式通过雷电接口，数据交换是在基于 PCIe 的商用计算机接口上完成的。

雷电接口的图标描绘的是闪电，而不是雷！商标是由英特尔公司注册的。

第二部分，符合 MPEG-2 规范的系统定义了两种多路复用方案。MPEG-2 程序流是一种复用方案，适用于单个程序存储或通过相对无差错介质进行传送；MPEG-2 传输流是另一种复用方案，适用于单个或多个程序存储或经由易出错的介质传送。这两种方案将在第 50 章进行概述。二者一般仅适于分发，而不适于视频采样或制作（索尼支持 SD 卡的 MPEG-2 IMX 方案是一个特例，它的 I-B 帧结构和相当高的数据传输速率使其适用于视频制作）。

17.4 IEEE 1394 (火线接口、i. LINK)

在 1995 年，IEEE 制定了通用高速串行总线的标准，该标准下的总线可以通过点对点的连接方式在一个树型网络中连接多达 63 个设备。该连接通过两条屏蔽双绞线 (shielded twisted pairs, STP) 传输数据，工作数据传输速率为 100、400、800Mb/s。每个点对点段的长度限制在 4.5m，整个网络横跨长度限制在 72m。提供异步和同步两种模式，后者可以接受实时数据。苹果使用它的商标火线 (FireWire) 来命名这种接口；索尼的商标是 i. LINK。

IEEE 1394 为高性能串行总线标准。

尽管没有实际运用到未压缩视频上，IEEE 1394 接口有足够的性能和前景运用在压缩视频流上，特别是数字视频上。

该接口使用标准 6 针脚接口为外围设备供电。索尼公司通常使用 4 针脚接口而没有遵从 IEEE 标准。一个节点可能是 4 针脚接口也可能是 6 针脚接口。4 针脚接口并不供电，许多人发现 4 针脚接口具有机械脆弱性。

国际电工技术委员会 (IEC) 为在 IEEE 1394 上传输的数字视频制定了标准。视频根据 BT. 601 标准进行数字化处理，再使用动态 MPEG 以 25M/s 的数据传输速率进行编码 (使用 DV 标准)，这也就是我们熟知的 1394/DV25 转换。高清数字视频编码率已达到 100Mb/s (DV100)，DV100 到 1394 之间的转换标准已被 IEC 采用。

这可参见 IEC 61883-1, 用户音频/视频设备-数字接口-第 1 部分：概述。也可以参见第 2 部分到第 5 部分。

17.5 HTTP 实时流媒体(HLS)

苹果公司已经推出了为万维网设计的 HTTP 协议的流媒体适配方案，该方案包括流传输和文件传输的各个方面。

HTTP：超文本传输协议。

服务器将节目分割成一系列的媒体文件，每个文件都含有很短的回放时间，通常是小的整数秒，如 10 s。通常情况下，某一分段为包含相同编码素材在不同码率下编码得到的媒体文件。每个媒体文件都是包含节目关联表（PAT）和节目映射表（PMT）的 MPEG-2 传输流，这就允许流解码时可以从任一分段开始，而不需要读取先前的分段。

这里的播放列表是扩展的 M3U 播放列表文件，可能是用 US-ASCII 编码的 .m3u 文件或者是用 UTF-8 编码的 .m3u8 文件。

该服务器还准备一个播放列表（索引）文件将时间偏移与一组媒体文件的文件名（即 URL），以及相应的数据传输速率和序列号关联起来。播放列表文件可以使用 gzip 形式进行压缩。

gzip：Gnu zip，一种无损数据压缩技术。

客户端建立一定的数据传输速率，期望数据传输速率维持不变，然后访问播放列表文件，确认期望时间和期望数据传输速率下内容对应的文件名。然后为该命名文件发送一个 HTTP 请求，再进行播放。

节目长度并没有预先设定好上界，故服务器可以随意添加附加的媒体文件，相应地更新（原子级）播放列表文件。要持续播放，就要在文件开始播放前的某个适当的时候（为估计延时作出规定），为播放列表中的下一个文件发送请求。客户端可以根据网络性能在不同的数据传输速率间动态切换。允许随机访问和“欺骗”式访问。

允许加入纯音频文件。规定是加密的。

这种方案称为实时流媒体。对于一个真正的实时“广播”，播放列表文件可以包含尽可能少的内容，如三个分段，包括根据加入时间和数据传输速率要求确定的客户访问分段。根据添加和删除媒体文件的需求，播放列表每隔几秒钟就会更新一次。在访问时会有一些延时，通常有几秒钟，比用 MPEG-2 的电视直播的延时要长得多。“直播”对于播放内容并没有要求：节目中的所有分段都会列在播放列表中，随着旧片段被保存，可以在节目播放中对播放列表进行追加。

所有的数据——无论是播放列表文件，还是媒体文件，都通过端口 80 使用普通网络协议传输：上行流使用 HTTP 协议获取播放列表文件的请求，像网页请求，下行流传输媒体文件就像传输庞大的网页。路由器、代理和防火墙的问题很少见。

HTTP 实时流技术运用在苹果的 Mac OS X 系统和 iPhone 手机上。

与其类似的一个方案，称为 HTTP 上动态自适应流技术（DASH），ISO/IEC MPEG 专家组正在为其制定标准。

第18章

元数据

本章在语气上与本书的其他章节会有所不同，我对元数据是持怀疑态度的。

元数据面临着诸多问题：机遇、商业活动、产品。然而就我看来，影视产业还没有达到足够的深度来理解元数据的深奥原理，这是别的方法无法达到的。

想想一个音频文件，存储着 2 亿对以 44.1kHz 采样的采样数据，用于呈现贝多芬第 9 交响曲欢乐颂。如想创作出接近原创的声音，首先需要了解采样率。

任何一本纸质文档中都会提供采样率，这差不多已成为一个标准。为了适应通用的解码器和播放器，编码采样率必须在文件中、或者文件头中清楚标明。

这是编码采样率数据还是元数据？

我认为这是数据，因为在未知的情况下，预先听觉经验是无法获得的。

前面音频文件的例子（这里称之为例 0）或许看起来有点牵强且不相干。那让我再举 5 个更深入的例子。例 1 在概念上比例 0 深入了一小步，我们将再由例 1 逐步进行到（专业视频复杂度和相关性也随之增长）例 5，其中涉及在视频工程中高度关注的问题。我认为例 5 与例 0 遇到相同的哲学难题：

什么是数据？什么是元数据？

由于这一难题仍然存在，那么以元数据为标题的这一章则需要提出问题，而不是提供解答了。

18.1 元数据例 1：CD-DA

CD-DA 是数字激光唱盘的缩写。CD-DA 是由飞利浦和索尼公司制定的，用于在容量约为 660MB 的光学媒介上，以 44.1kHz 采样率、16 位采样（即约 1.5Mb/s 的数据传输速率）存储 Hi-Fi 级数字立体音频。

CD-DA 是由索尼和飞利浦的“红皮书”定义的，随后 IEC 作为 IEC 60908 标准化。

原来的 CD-DA “红皮书” 中没有包括专辑主打、艺术家、唱片说明文字和其他文字信息的任何规范。这些信息印在 CD 封套上。显然飞利浦和索尼认为用数字格式提供这些信息是多余的！CD 格式不但缺少元数据，还缺少独特的 ID。

几年后，CD 的支持者采用了 CD 文本标准，补充了红皮书使其允许记录基于文本的元数据。但那时已经太晚了。

刻录的 CD-DA 媒介必须包括提供的轨道数、轨道起始时间、轨道持续时间（精度为 1/75 s）内容的列表。身兼音乐发烧友和软件工程师的 Ti Kan 发现这些信息可以编码成 32 位的数字，将其代替唯一的 ID。当 CD 开始流行时，Kan（在 Steve Scherf 的帮助下）建立了光碟资料库（CDDB），这是一个数据库，用于存储由大众提供的与其代码相关的元数据。CDDB 最初是为大众主导服务的，但随后变成了商业实体：一开始是 CDDDB 公司（1995 年），之后是 Gracenote 公司（2000 年，随后在 2008 年被索尼收购）。

今天有些人会称列表的内容为技术元数据。我认为这就是数据：如果没有节目列表，用户无法让系统发挥预定作用：播放歌曲。

因此，CD 唱片拥有元数据，但不是可靠地来源于或者直接受控于内容创作者。

这个教训对于系统设计者是这样的：什么构成“数据”，以及什么构成元数据是根据你对系统界限的界定来划分的。索尼和飞利浦显然将 CD 系统认为是分布式预录制的数字音频系统。今天，我们认为 CD 系统是给用户发布音乐的。这里一个细小的差异改变了什么是数据，什么是元数据。

当 MP3 音频压缩系统被开发出来时，开发者规定了 ID3 标签来说明元数据来源于内容创作者。

通常用于广播音频的 BWF 文件格式包含一个“参数”叫做 nSamplesPerSec，用于赋予采样率。该参数包含在“BWF 元数据块”中。那么采样率是元数据吗？

见 ITU-R BR. 1352-2 中的“广播波形格式”(BWF)。

18.2 元数据例 2：.yuv 文件

.yuv 文件格式是 Abekas 在 19 世纪 80 年代末提出的，用于存储未压缩的视频文件。例如，对 8 位的 $Y'C_bC_r$ 色彩空间，在光栅顺序为 4 : 2 : 2 的隔行视频中，这种文件格式定义基本如下：

存储连续的图像行，每一行都是一系列 4 字节的元素 $(C_{B0}, Y'_0, C_{R0}, Y'_1)$ ，其中下标“0”表示偶数的亮度采样，下标“1”表示奇数的亮度采样。

“.yuv”文件没有头文件，而且也没有存储帧数、图片的行数和图片的列数。这种格式曾推荐用于存储分辨率为 720×480 像素的视频，后来也应用到分辨率为 720×576 像素的视频，也可应用于分辨率为 720×481 、 720×483 、 720×486 像素或 704×480 像素的视频中，还被应用到分辨率为 $1280 \times 720p$ 及 $1920 \times 1080i$ 的解码研究中。

考虑到.yuv 文件的读取被限制在 720×480 像素或 720×576 像素的分辨率上。大多数时候可以将文件位数除以 1440，然后依次除以 480 和 576，看除以哪个数得到的商是整数，从而来确定格式。但这种方法有时并不可行，比如，一个 4,147,200 位的文件既可以是 480i 的 6 帧文件，也可以是 576 的 5 帧文件。

可靠的文件解释只能通过发送者和接收者之间的协议来实现，或更准确地描述为作者和读者之间的文件条款，也即属于文件自身传输之外的范畴。

试想，通过预加含有 3 个 32 位字的文件头来扩展.yuv 文件格式，三个位分别是：帧数的计数、图像行数的计数、图像列数的计数。这是头数据还是元数据？如果你的“系统”预先定义为 480i 的，那么文件头中的计数就是没用的辅助信息——称为元数据。但是如果你的“系统”是多格式，那么计数是必要的数据，因为该文件的图像部分的可靠解释没有文件头中的计数是不可能实现的。

得到的结论如下：什么样的系统包含“元数据”取决于你对你的“系统”的定位。你的系统越大，包容性越强，越具有一般性，你对内容的依赖性越小，则你系统中越多的元数据就会变成数据。

18.3 元数据例 3：RFF

自 1953 年来，电视节目的主要来源一直是电影——首先是胶片电影，然后是数字电影。在超过半个世纪的时间里，电影都是以 24Hz 的帧频播放的。北美电视的早先的场速率为 59.94Hz，为匹配电影的帧频，较好的解决方法为，将电影的帧频降到 23.976Hz，然后加入 2-3 下拉模式，通过将连续的电影帧画面播放 2 次，播放 3 次，再播放 2 次，播放 3 次，如此循环。由此造成的一定程度的运动停帧，并不会引起用户不满。某些视频的

帧——例如 M 帧，详见图 34-1——由来自两个不同电影帧的字段组成。

大约在 1990 年左右，用户电视的接收机已经可以通过去隔行扫描（借用数字化手段）和以 59.94 Hz 的速率播放帧来消除播放抖动的缺陷。由于普遍使用“电影”视频材料，去隔行扫描需要对 M 帧进行检测和处理。

这里采用的去隔行扫描技术对连续视频场中的图像数据元素进行比较，查看是否识别到 2-3 下拉模式。如果检测到一个持续的 2-3 下拉序列，则认为该视频源为 24 Hz，相应地重新组装视频帧。随着 CE 技术的进步，接收机变得越来越依赖这样的算法，现在一个高品质的数字电视处理器芯片完成这一任务需要提供 10 万个门电路。问题是实现方法不一定可靠，不同的实现方法又不一致。

当广播“21 行”隐藏字幕数据变得司空见惯，每场大约传送 16 位的数据时，问题就会凸显出来。2-3 下拉问题可通过在电影每场中添加 1 位信号标识电影的下拉，将其消灭在萌芽状态。

MPEG-2 系统通过在图像编码扩展中传送重复第一个字段（Repeat First Field, RFF）标志来兼容 24 Hz 的视频。该标志导致该字段对的第一个字段被重复。MPEG-2 的 REF 标志可以认为是元数据的“暗示”：忽略它可以获得满意的性能，但是使用它可以提升性能。

18.4 元数据例 4：JPEG 或 JFIF

在成立之初，JPEG 委员会决定避免色域的冲突：将其范围确定为压缩和解压缩图像数据，而不关心数据代表的是什么颜色。JPEG 可以为单通道灰度图像数据、三通道图像数据，例如 RGB，以及四通道图像数据，例如 CMYK。

色度下采样被认为是提供了一个大的压缩增益——在 4:2:0 的情况下能达到 2:1——所以允许亮度-色度编码作为一个预处理步骤。

JFIF 要求 BT. 601 的 $Y' C_B C_R$ 进行亮度-色度矩阵变换；但是 JFIF 的 4:2:0 色度二次采样是在水平和垂直两个方向上选址的，这和 BT. 601 不同。

在万维网兴起之前，JPEG 技术已经发展到了顶峰——JPEG 格式的原始目标应用是彩色的传真！格式的制定者希望系统集成商为非 JPEG 格式的色彩可靠传递制定规范。然而 JPEG 迅速被作为文件中彩色图像交换的方法来使用，而不仅仅是用于压缩图像。

斯高柏 (C Cube)，很显然作为 JPEG 的支持者，对于 JPEG 文件交换中色域混乱的问题，扬言要抑制商业化。斯高柏很快起草了一份文件（文件元数据）定义了一种称为 JFIF 的文件格式，声明图像数据在 BT. 601 的 $Y' C_B C_R$ 色彩空间中进行编码（没有上余量和下余量），JFIF 在算法上是明确的。然而，JFIF 对于 C_B/C_R 的参考范围要求为 ±128，根据这一范围来实施，也就是 257 个整数——但是 8 位编码最多只有 256 个数！为了满足生产规范，灰色有了非整数的 C_B 和 C_R 值，灰色的解码也注定是有色的。实施方式（特别是广泛使用的 libjpeg 库）使用标准化的比例系数，但将 C_B 和 C_R 的值限制到 +127，会对纯蓝和纯红色有所裁剪。

Hamilton, Eric (1992), JPEG 文件交换格式，版本 1.02（加州 Milpitas 的斯高柏微处理系统）这种非正式文档被 ECMA 认可，并略做改动，在 2009 年 6 月出版的 ECMA TR/98 上刊出相同的标题。

JFIF 明确规定了 BT. 601 的亮度-色度矩阵，而 BT. 601 没有提到基色的内容，JFIF 也没有提到打算用哪种基色。JFIF 被计算机工业广泛接受时，BT. 709 基色正用于制定 sRGB 标准。在实际中，JFIF 使用 BT. 709 基色。

为将 JFIF 用于 BT. 709 的高清 $Y' C_B C_R$ ，且符合标准则必须使用⁶⁰¹ $Y' C_B C_R$ 编码、压

缩、传输、解压缩，最后重新编码为⁷⁰⁹Y'C_BC_R。

JFIF 规范规定，由 YC_BC_R 线性转换计算 RGB 分量时不得进行伽马校正（伽马值 = 1.0）。这段话并不是说线性光分量不需要编码；而是说解码是为了符合 BT. 601 的做法，（在求得由 Y'C_BC_R 到 R'G'B' 的去矩阵化之后）把一个 2.4 次幂函数加到 RGB 分量上来显示三刺激值。

所以，就遇到了下面的难题：

- 该规范提到“BT. 601 的 Y'C_BC_R”，但是却违背了 BT. 601 标准。并且没有任何理由就使用了“全摆幅”。
- 该规范指出 C_R 和 C_B 值的范围为 ±128，已经超出了 8 位整数的运算范围，必须通过裁剪纯蓝色和纯红色来实现。
- 该规范提到“伽马值 = 1.0”，但其目的显然不是线性光编码。该规范并不使用伽马校正，而是隐含了一个 2.4 次幂律的电光转换函数（EOCF）。
- 该规范提到“BT. 601”，显然用它来定义基色，但 BT. 601 并没有提到基色。在实际上使用的是 BT. 709 的基色。

你可能会认为，“这只是一个关于元数据没写好的规范和未按照规范执行的故事。”这个评价基本上是正确的。下一个例子与之很相似，不过是将未设计好的元数据加入到了专业的视频流之中。

18.5 元数据例 5：序列显示扩展

与 MPEG-2 的 RFF 标志某种程度上近似一致，MPEG-2 的序列显示扩展提供了有关图像信息如何显示的解码器。该标准提供了（以印刷形式，作为元数据）一些列表，表中给出了 RGB 基色色度（颜色的基色）、传递函数（传递特性），以及亮度-色度矩阵（矩阵系数）。位流将这些枚举代码作为索引传送到列表中。H. 264 采用了相同的方案。这些列表总结在表 18-1、表 18-2 和表 18-3 中。表 18-1 代码 4 和代码 7 已经被弃用；NTSC 1953 的条目已经完全被弃用。代码 5 和代码 6 不适合高清（HD）。代码 8 在“一般的电影”中是无法预测的。无论在位流编码时使用哪种代码，基色肯定由 BT. 709 来提供（粗体）。表 18-2 中 MEPG 提供显示伽马值，读作 EOCF；这些条目用 † 加以标记。所有其他的条目定义为 OECF。代码 8、代码 9 和代码 10 是无法实现的。代码 1 和代码 6 具有相同的解释。表 18-3 代码 4、代码 5 和代码 6 有相同的解释。GBR 在 H. 264 的 VUI 中，用于 R'G'B' 4 : 4 : 4 编码。

这里描述的是 MPEG-2 的序列显示扩展，相同的元数据在 H. 264 的附录 E，视频可用性信息（VUI）中传达。

基色码 4 指定了 NTSC 1953 的基色。据我所知，没有任何现存录制的视频材料使用这些基色，在引进磁带录像机之前它们已经被弃用了。位流中包含该值是荒谬的：如果遇到了这段代码，所有的解码器都应将这段忽略掉（好心的技术人员可能为已设置的代码着想，“我在广播 NTSC 制式广播，这是唯一可以声明其制式为 NTSC 的设置，最好使用它”）。

见 DeMarsh, LeRoy E. (1993) 的“电视显示荧光粉或基色：有些历史”，SMPTE J. 102 (12): 1095-1098。

对于传递函数（表 18-2），MPEG-2 无法区分 OECF 和 EOCF，表中包含二者。两种代码都可用于 BT. 709。代码 5，显示伽马值为 2.8，从来没有在实际中使用过，即使是在欧洲（其他标准提到过伽马值 2.8）。线性传递函数，代码 8，当每个分量使用少于 14 位

的编码时得到的图像质量是不可接受的。对数编码是不可行的：这样的编码会对低照度颜色进行裁剪，因而在一定程度上会引起消费者的不满。

表 18-1 色彩基色

代 码	解 释
0	已禁用
1	BT. 709
2	未指定
3	已保留/用于将来
4	BT. 470-6/NTSC 1953
5	EBU Tech. 3213
6	SMPTE EP 145
7	SMPTE 240M
8	“Generic film”

表 18-2 传输特性

代码	解 释
0	已禁用
1	BT. 709
2	未指定
3	已保留/用于将来
4	显示伽马值 2.2
5	显示伽马值 2.8
6	BT. 709
7	SMPTE 240M
8	Linear
9	Log ($10^2 : 1$)
10	Log ($10^{2.5} : 1$)
11	xvYCC
12	BT. 1361
13, ..., 255	已保留

表 18-3 矩阵系数

代 码	解 释
0	已禁用
1	BT. 709
2	未指定
3	已保留/用于将来
4	BT. 601
5	BT. 601
6	BT. 601
7	SMPTE 240M
8	$Y' C_G C_O$
9, ..., 255	已保留

MPEG 和 H. 264 具有普遍概念模型，只有位流和解码器是标准化的。一个兼容的编码器只发送合乎规定的位流。此外，编码器的其他方面都不是标准化的，所以传输特性应该定义为 EOCF，而不是 OECF！

至于亮度-色度矩阵（表 18-3），BT. 601 的设置是三重的，没有很好的理由。

这些表格体现了之前所说的百科全书式得到元数据：把所有可能性都收集进去而不考虑实际使用情况，不提供有关如何编码和解码元数据的指导。

空间数据元数据相关的问题应该由软件开发商来解决。以下是苹果公司关于传输特性中的双重 BT. 709 代码的叙述：

QuickTime 的编写者应该在传输特性转换的时候将代码 6 映射到代码 1 上……

MPEG-2 规范引用了 ITU-R 的 Rec. BT. 470-6。关于这些引用，苹果公司写到：

这些信息是不完整且过时的。

应该再加上“错误的”。

MPEG 错误地引用了 ITU-R 的文档：被引用的文档是一份报告，而不是建议。

在这种混乱的情况下，编码器应如何配置？

- 对于标清视频，将 576i 的文件基色设为 EBU 3213，将 480i 的文件基色设为 SMPTE RP 145。将传输特性设为 BT. 709，矩阵系数设为 BT. 601。对于高清来说，在 SDE 中到处都声明了 BT. 709，ATSC 编码器的一个问题是在北美高清视频都使用 SMPTE RP 145 基色，且你尝试声明 SMPTE RP 145 基色；然而 ATSC 规范中指定了 BT. 709 基色，而且所有的用户接收显示都是以 BT. 709 为基色的。我的建议是声明 BT. 709，原因有两点：要遵循 ATSC，当区域基色设置成为历史时为未来做好准备。解码器需要做什么？
- 对于标清格式，如果将 BT. 709 定为色彩基色，则需要对它加以重视，否则，期望将 EBU 3213 作为 576i 文件的基色，将 SMPTE RP 145 作为 480i 文件的基色，其

他的代码都是无意义的，都应看成 BT. 709。将传输特性设为 BT. 709，而不用管有没有声明其他基色。除了将矩阵系数设为 BT. 601 之外，BT. 709 应该是正确的，不过也要保持怀疑的态度。其他的代码几乎肯定是错的。

- 对于高清格式，全面承认 BT. 709，对其他代码保持高度怀疑。

对编码器处理的建议总结在表 18-4、表 18-5 和表 18-6 中。

表 18-4 色彩基色的解释

代 码	解 释	代 码	解 释
0	已禁用	5	用于 SD
1	用于 HD 或 SD	6	用于 SD
2	未指定	7	使用代码 1
3	已保留/用于将来	8	使用代码 1
4	使用代码 1		

表 18-5 传输特性的解释

代 码	解 释	代 码	解 释
0	已禁用	7	使用代码 1
1	使用	8	使用代码 1
2	未指定	9	使用代码 1
3	已保留/用于将来	10	使用代码 1
4	使用代码 1	11	xvYCC：实现
5	使用代码 1	12	使用代码 1
6	用代码 1 代替	13, …, 255	已保留

注：代码 11 xvYCC 有可能会成为未来的系统，但是现在不太可能实现。

表 18-6 矩阵系数解释

代 码	解 释	代 码	解 释
0	已禁用	5	用于 SD
1	用于 HD	6	用于 SD
2	未指定	7	使用代码 1
3	已保留/用于将来	8	$Y' C_G C_0$ ：实现
4	用于 SD	9, …, 255	已保留

注：代码 8 ($Y' C_G C_0$) 可能会用于未来的系统。我不知道现在的解码器能不能解码 $Y' C_G C_0$ 的色彩空间，不过明显没有好处。现在来说也极不可能实现。

18.6 小结

在演播室中，MXF 在文件中的元数据能可靠地编码、存储和传输。ANC 数据包通过 HD-SDI 数据流传送元数据；不过，涉及如何处理随着视频数据通过处理设备的元数据，并不存在什么常见、或广泛使用的方案。

除了 MXF 文件和 HD-SDI ANC 元数据之外，制片设备没有可靠的元数据系统。

基于文件的工作流程能保证可靠的元数据处理。然而，元数据的概念问题（如例 0 到例 5 所示）依然存在。对此提出以下建议来实现元数据系统：

- 精心设计元数据。保证元数据编码以及在解码时的解释元数据的规则清楚。但不要面面俱到：“让系统越简单越好，但不要太简单”（爱因斯坦）。

- 元数据应该是无关紧要的，否则就会成为数据了：设计的系统在没有元数据时能够工作，加上元数据来增强功能。
- 力求一种设计使得添加元数据后不会损害基本系统运行。
- 如果信息已经在数据里了，不需要将其复制进元数据：这样做会造成（在稍后的时间里）信息冲突的可能性。
- 如果在一个特定的系统设计里，可以从根本上分离元数据，基本格式的选择必须保证在缺少元数据情况下能够为自身的目的服务。

第19章

立体（三维）视频

立体三维（Stereoscopic 3D, S3D）是指以两个视角，采样、处理、存储、分配，并显示图像，一个视角用于左眼，一个视角用于右眼。观测画面通常通过摄像机在横向距离很小的不同位置观测同一场景来获得。立体视觉呈现一种错觉，不同于观察现实世界，观看的景色不变，只是观察者移动了他或她的头。尽管如此，对于那些精心准备的素材，效果还是令人信服的，在某些情况下，可以添加到故事素材中。

术语 S3D 将图像深度信号（特别是角度）与立体三维区分开，但只有一个图像。计算机生成图像（CGI）是从场景中综合生成的图像，CGI 相对来说比较简单。有些人认为术语 S3D 是多余的——立体一定是三维的。

19.1 图像获取

为了以专业的内容创作的角度，从真实场景中获取图像，通常需要两台摄像机，每一台都需要成像器和信号处理装置。为了生成“正常的”立体图像，摄像机的光轴摆放的距离要等同于人双眼分开的距离——也即双眼间的距离（俗称瞳孔间距），成人的瞳孔间距约为 52mm~75mm，平均距离约 63.5mm (2.5in)。不同的显示效果可以通过调整摄像机轴间距离来做到：设置一个较宽的摄像机轴间距离并减小深度，这样就会使场景比实际要小；设置一个较窄的摄像机轴间距离并扩大深度，这样就会使场景比实际要大。失调的摄像头会导致观察者观感不适。

经常会使用两个照相机，不过，许多其他的配置也已经证明可用，例如一个镜头两个成像仪和两个镜头一个成像仪。

19.2 S3D 显示

通过为每只眼睛进行专业显示，能以历史观影大师的方式呈现 S3D。从 20 世纪 90 年代到 21 世纪，许多虚拟现实系统使用这项技术，有时和头部追踪设备组合使用；但是用户感觉头戴显示设备并不舒适！保持一定距离观看是商业需求。

正常电视的观看距离约为 3m，对这一观看距离，有多种方案可以用于将两路影像在显示设备上复用，以及将影像为观影者的两只眼睛进行分离：立体电影、时间多路复用、偏振、波长复用、视差屏障裸眼三维和光栅裸眼三维显示，这些技术会在后续几节中进行简述。

将要介绍的这些技术几乎总是使用单一的“本机”二维显示（直接观看或者是投影）。在这种情况下，所有的技术都有一个缺点，即至多只有 50% 的本机二维显示的光是可用的（经常会更少），所以三维显示系统会比较暗。

19.3 立体电影

图像创建的方法是将左侧画面的红色分量作为红基色，以及右侧画面的绿色和蓝色分量作为原本是二维视频流的三个分量（显然，这些假设需要进行色度的亚采样，而 H. 264

或 MPEG 编码就会被破坏)。

将红色与左侧关联，来符合航海规则中红色代表端口左侧的规定。

显示时用红基色呈现左眼的图像，用绿色和蓝色呈现右眼的图像。

观看者的双眼都佩戴着具有滤色功能的眼镜，红色滤光片放在左眼上——左眼只能看到信号的红基色，包含整个左侧画面。蓝绿色的滤光片放置在右眼上，右眼可以看到右侧画面中绿色分量和蓝色分量的两色组合。全彩色相对于每只眼睛、每个像素点来说都是不存在的，尽管如此，观看者的视觉系统的损失也得到了很大程度的补偿（尽管会有一些不舒服）。红色/蓝绿色方案是最常见的，不过立体电影显示也可以使用其他的颜色组合。

由于使用标准的二维视频基础设施便于获取和传输（诚然，其通常的假设之外），立体电影方案在电影院和电视上零星地使用了数年，大多都被淘汰了，现在通常被认为是一时新鲜。

19.4 时间多路复用

两路画面可以在时间上复用，显示操作是以（至少）两倍帧频交替地呈现左眼画面和右眼画面。

观看者佩戴主动式快门眼镜，并同步显示，这样左眼被挡上时会显示右边的画面，当右眼被挡上时会显示左边画面。

快门同步通常是通过以帧频发射红外线（IR）光束，掩盖可视区域来实现的。每一副眼镜都有一个红外接收器（现在已提出蓝牙射频同步）。

该方案在三维消费电视中占主导地位，在电影院（XPAND 3D）中使用有限。

19.5 偏光法

许多 S3D 显示方案设计涉及偏振光。最简单的极化方法——用于商业的线性偏振（linear polarization, LP）和圆偏振（circular polarization, CP）。观看者佩戴被动式偏光眼镜，双眼的滤光器会有相反的偏光效果。

物理上关于偏振光很好的阐述见莱因哈德，埃里克等（2008 年）著的书：《彩色成像：基本原理与应用》（韦尔斯利，马萨诸塞州：AK 彼得斯）。

偏光法也可以是时间复用的：显示器在（至少）2 倍图像帧频下工作，并交替显示左眼画面（以一种偏光方式）和右眼画面（以相反的偏光方式）。

在影院中常使用的 RealD 系统中，在放映机的光路上，放映镜头和端口玻璃之间插入一个“Z 屏”，Z 屏是一种电光设备，可以迅速切换圆偏振的极性。成像器按时序生成左眼图像和右眼图像；Z 屏也同步驱动（正如之前提到的影院中配置的 RealD 系统，每（1/24）s 有三个左右循环——也就是说，显示调制器以 144Hz 的频率生成图像）。该技术因为直接显示图像而无法商业化。

偏光放映可以在同一时间生成两种图像——例如使用一对放映机（或者两个图像调制器共享同一个放映镜头）。不过，这种解决方案没有得到广泛应用，这归因于其高昂的费用。单个 4K (4096×2160 像素) 放映机可在顶部显示一个 2K (2048×1060 像素) 的左侧图像，并在底部显示一个相同大小的右侧图像，然后配置光学器件以对这两个图像进行相对偏振极化，并将其组合起来同步显示。索尼公司已将这项技术在影院中商业化使用了。

作为一个商业化的系统，每幅图像有 1060 行，而不是我们以为的 1080 行：两幅图像之间有 40 行消隐行。

偏光放映要求屏幕一直保持着偏振极化状态。典型电影荧屏去极化，常使用“银制”屏幕，其实是镀铝的屏幕。

为直观显示，偏振可以通过为显示中交替出现的图像行制造极性相反的偏振片来完成。显然，在三维显示操作中，垂直分辨率相比较本机显示能力来说会减半。这样的显示可以用于呈现普通二维影像而无需眼镜（尽管至多只有 50% 的二维显示光线是可用的）。

交替出现的相反极化图像行常通过使用薄膜式偏光眼镜（film pattern retarder, FPR）来实现。

偏光系统的一大优点是，眼镜是无源的而且也不贵。

19.6 波长复用（光谱立体/杜比）

这项技术是由德国戴姆勒-奔驰公司的 Helmut Jorke 发明的。显示器以 2 倍（或更高）图像帧频工作，然后透过不同的光学滤片，首先呈现左眼的画面，接着是右眼的画面。每一对波长组合（例如 G_{LEFT} 和 G_{RIGHT} ）被设计成几乎不重叠。该光学滤波器的特性通过信号处理进行补偿以产生大致的同色异谱对，即波长的红色部分构成不同，但颜色看上去大致相同。

观看者佩戴被动式镜片，每块镜片上都有与放映机大致相匹配的光学滤片，左眼的滤片滤掉了 R_{RIGHT} 、 G_{RIGHT} 和 B_{RIGHT} 相对应的波长部分，右眼滤掉了 R_{LEFT} 、 G_{LEFT} 和 B_{LEFT} 对应的波长部分。

波长复用方式可以同时呈现左图像和右图像。不过，这一模式还没有商业化。

光谱立体方案使用被动式眼镜（尽管有些昂贵），但它不需要保持偏振极化的屏幕。

杜比将该方案投入商业化用于三维影院，但还没有商业化的直观视图显示设备。

19.7 裸眼立体显示

裸眼立体指的是呈现立体图像但却不需要观看者佩戴眼镜的技术。其中有两项技术已获得有限的商业化：视差屏障技术和光栅技术。

裸眼立体显示通常会在一个体积小的视听空间中创造一个合理的立体声系统。主要问题在于“最佳点”通常相对较小，而在最佳点之外，立体声效果会显著降低，甚至完全消失。另外，裸眼显示有时会（意外地）产生观看图像是相反的观看点而导致深度反演的现象，也就是我们熟知的伪立体。

19.8 视差屏障显示

两个图像在同一显示平面上交错逐列显示，在显示屏前面一小段距离有一组屏障构成的狭缝，在正常的观看距离下，可以看到交替的图像列。屏障的几何形状（俯仰角度和位置）被设计为处在选定的最佳观看位置，一组列相对于左眼是可见的，另外一组由于屏障是“阴影部分”，相对于右眼，情况正好相反。

这项技术已经商业应用于手持设备（三维相机和手机）。

19.9 立体光栅显示

两个或多个（ N 个）图像在显示器上以 N 列为一组交错扫描，显示器前放置着一组光栅，每 N 列放置一个微光栅。每一组镜头的几何结构被布置成：在显示器前面的空间可以交替放映多个图像的列。如果只有两个图像（ $N=2$ ），左右图像就会交替播放。

飞利浦公司展示了光栅裸眼立体显示可以通过基于单个二维图像及其伴随的“深度地

图”(2D+Z)的信号处理在显示器上显示多个图像($N \approx 9$)，“深度地图”是在图像的后期制造时进行编码的，或图形产生的(例如在计算机游戏中)。该技术在数字标牌中已有有限的使用，但没有将其商业化到消费应用中。

深度图可以轻易地制作出CGI内容，包括用户的计算机游戏。然而，不存在将深度图从计算机传送到显示器的广泛可用的标准。深度图技术无法直接处理阻塞问题，所以其视觉性能是有限的。

19.10 录制和压缩

对于一个给定的图像格式(例如 1920×1080 像素)，S3D通过将数据传输速率提高为单一图像数据传输速率2倍的一对图像来获得。传输和接口的关键问题为高速率。专业的获取和后期制作需要2倍数据传输速率(通常需要两倍的专业设备)。为了便于用户录制和发行，已经设计出小于2倍二维图像数据传输速率的三维系统。

已经设计了很多技术用来录制S3D的内容，并将其内容通过广播分布链进行传输。一些分布网络将左图像和右图像以2:1的比例挤压在一起，将水平相邻的方式组合为单一信号，从而可以在普通的分布网络中传送。

蓝光标准中增加了一种采用H.264协议中的立体高配压缩S3D内容的机制。H.264的运动估计和运动补偿差值方案被设计用来对具有高度的空间相关性的图像进行简单序列编码，而图像的差值是由曝光时运行时间不同引起的。立体图像对的左图像和右图像表现出高度的空间相关性，图像间的差值来自于由视差引起的位置偏移(差距)。在典型的SHP应用中，右图像是左图像用差值向量(相当于运动向量)进行“运动”补偿后加以预测而得到的。通常立体图像可以二维图像数据传输速率的1.3倍或1.6倍进行编码。

立体声高配与H.264的多视角配置(MVP)有关：二者都记录在当前版本的附录H中。有些人可能会引用低至1.2或高达1.8的倍数。

19.11 用户接口和显示

上节已经讨论了S3D图像的获取和显示，本节将讨论用户显示接口。

HDMI 1.4a版有一个强制性框架打包三维结构，其将左右 1920×1080 像素图像打包成 1920×2205 像素的“容器”，其中包含45条消隐线用于分离图像。并有逐行和隔行两个版本。

Dave LeHoty将他的家庭高清系统描述为1.3a加一个星号，暗指各种各样的版本和选项，这使得系统集成对专业人士来说很困难，更不用说对于普通用户了。

HDMI 1.4a版还描述了用 1920×1080 像素容器传送 960×1080 像素左眼图像和 960×1080 像素的右眼图像的接口，水平方向上都按2:1进行压缩，相邻并排(SbS)放置，水平分辨率减小。

最后，HDMI 1.4a版还描述了一个用 1920×1080 像素的容器来传送 1920×540 像素的左眼视图和 1920×540 像素的右眼视图的接口，垂直方向上均按2:1进行压缩，按从顶到底(TaB)的顺序放置，垂直分辨率减小。

我曾经探访过一个消费类电子产品零售商，该零售商正通过HDMI线从蓝光光盘上顺序传送三维电影——但是显示设备上三维效果被禁用。我调整我的聚散度，希望可以裸眼观看三维图像。一个销售员上前说道：“这不是三维。”我说：“我只是在看其深度。”不带一丝怀疑，他充满自信地说：“不，你不是。”

还有许多方案，比较混乱。

19.12 幻影

大部分本章所描述的显示技术遇到的问题是，本应是为左眼显示的光却“漏到”右眼，或反之。可以称之为“串扰”。这些由不需要的光创造出来的人工图像称为幻影。

产生幻影有多种原因。在多数显示器中，针对视频信号产生的光不是瞬时产生的。例如，液晶显示器的液晶屏需要一定时间对驱动信号做出响应；等离子体显示面板的荧光体也会有一定的延迟时间。当液晶显示器和等离子体显示面板用作三维图像显示时，采用时间复用方式，当相反的快门打开时显示器还有延迟，幻影就会出现。在偏光显示器中，偏光片（在显示器和眼镜上）通常会消光不完全。在光谱立体方案中，实际的光学滤波器会出现一定程度上的不希望出现的频谱重叠。

要将幻影降低到可容忍的程度，需要在其显示之前进行补偿（在影院，这个过程称为幻影消除 ghost busting）。如果预先知道左侧明亮的光会泄露到右方，可以人为地去除右侧对应位置的图像。不过补偿必然是不完善的，如果右侧图像对应位置是黑色的，没有光可以减去，但是串扰依然存在。在电影院，无论是母带还是放映机的信号处理，都可以实现补偿。电影制作者不希望为每种三维显示技术制作不同的母带，所以第二种选择比较常见。

19.13 聚散度和调节

人类视网膜由视轴相交的区域是凹的，它由一群紧凑的感光细胞组成。视网膜中心凹角直径约为 1° ，它涵盖一小部分视觉区域——相对于在正常观看距离观看高清图像的千分之几的区域。

眼睛的动眼系统包括眼球周围的肌肉。这些肌肉“控制”着每只眼睛上的视轴使得中心凹区域的光线聚焦在视觉区域上。每秒会有几次肌肉操作和视线转移到一个新的点，该动作称为扫视。

在实际场景的双眼视觉中，当双眼的视轴遇到感兴趣的场景元素时，双眼就会随之做出动作。动眼系统对视轴相交的距离的控制称为聚散度。

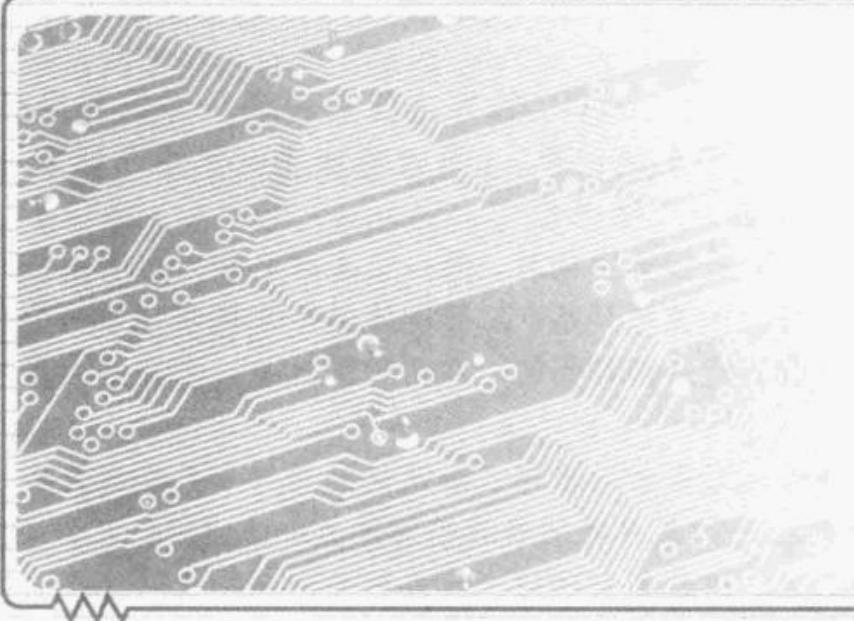
聚散度运动理想情况下是眼球关于其平面中心转动。

人的晶状体密封在一个胶囊体中，具有一定的柔韧性：晶状体可以改变形状。在眼球里，晶状体的周围有一种肌肉——睫状肌。当肌肉处于松弛状态时，晶状体囊是平坦的，晶状体的焦距为最大。当睫状肌收缩时，晶状体囊变得更像球形，焦距减小，侧重于较近的物体。晶状体上的肌肉控制称为调节；类似于相机镜头的对焦。

老花眼是由于年龄，晶状体变得不容易弯曲，造成的适应能力下降，即使是有正常视力的人，超过 50 岁后也需要戴眼镜。

通常人类视觉观察真实物体时，聚散度和调节系统协调工作。

在之前所描述的各种立体显示中，左右视图都会在显示器上显示出来，显示器表面到观看者的距离是固定的。为了保持图像的敏锐度，需要调节观看者到显示屏的距离，而不是立体显示系统设计的与观看对象间的观测距离。当视觉深度偏离屏幕距离时，无论是一个远距（在屏幕之“后”）还是一个近距（在屏幕之“前”），观看者由于观看经验很容易感受到聚散度和调节之间的冲突（V-A 冲突）。研究人员已经证明，V-A 冲突是观看三维立体时观看者感到不适的主要原因。



第二部分

理 论

第 20 章

滤波和采样

本章主要解释一维信号在模/数 (A/D) 转换之前如何进行滤波和采样，以及如何在接下来的数/模 (D/A) 转换中重建。第 21 章会将这些内容扩展到数字域的转换中。本书第 22 章将这些概念扩展到二维图像中。

当一维信号（例如音频信号）进行数字化时，每一个采样必须以单个值进行封装，在采样过程中可能是以一个复杂的模拟波形开始。二维图像进行采样时，每一个采样可能是以图像平面内小部分区域的复杂能量分布开始的。在每种情况下，潜在的大量信息都必须缩减到单个数字。

本书之前描述了对模拟信号波形最初的采样。如果读者觉得在数字域更合适，考虑用因子 n （如， $n=16$ ）对图像的行进行采样收缩以调整图像大小的问题，且需要对每组 n 个输入采样点都计算一个输出采样。这就是数字域中重采样的问题。它的约束与对原始模拟信号进行原始采样的约束相同。

在采样前，必须舍弃掉采样间隔中的细节。在采样前减少信息的过程是预滤波。采样的挑战在于丢弃信息时，避免大于采样间隔的尺度上的信息丢失，在采样的整个时段避免引入失真。采样理论阐述了信号可以进行采样和准确重建，以及信号会受到一些不可避免的细节损失，这种情况下，信号无法用一定数量的采样点在数字域加以表示。

采样理论最初用于描述一维信号，如音频，其中的信号是单一时间维的连续函数。采样理论已经扩展到图像上，其中图像被认为是二维空间坐标系（水平和垂直）中的连续函数。采样理论可以进一步扩展为运动图像的时间采样，这里的第三维是时间维。

20.1 采样定理

假设一个信号进行了很好的数字化处理，函数在时间上变化很小，考虑如图 20-1 所示的余弦信号，其中 x 轴表示采样间隔，图中上方的波形为余弦函数，其频率为采样频率 f_s 的 0.35 倍，中间的波形频率为 $0.65f_s$ 。使用的符号约定如下： $\omega = 2\pi f_s$ 单位为 $(\text{rad} \cdot \text{s}^{-1})$ 、 $t_s = 1/f_s$ 。图中分别采用了采样频率的 0.35 和 0.65 频率，下方的图表示对于这两个波形采样得到了相同的采样结果：两个波形可以得到相同的采样序列。如果对中间的波形进行采样，然后按常规方法重建，则会得到上面的波形。这就是混叠现象。

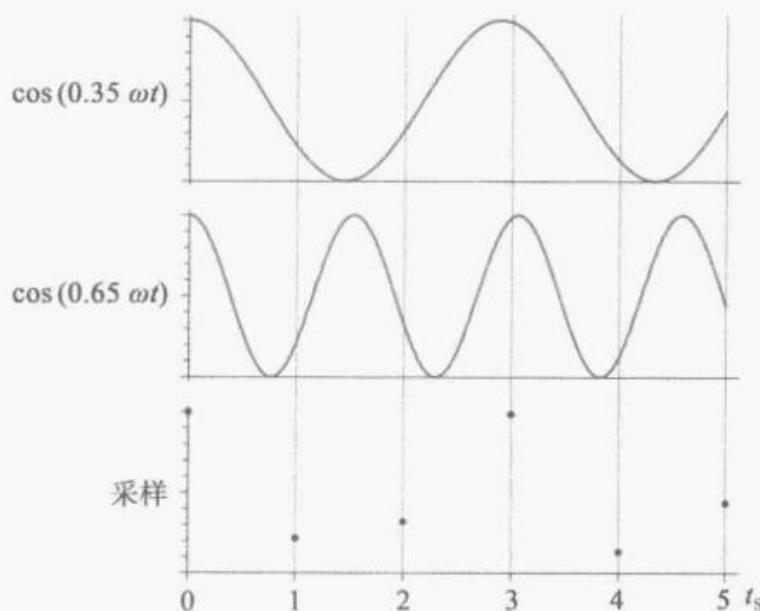


图 20-1 频率小于或大于 $0.5f_s$ 的余弦波

20.2 对频率为 $0.5f_s$ 的信号进行采样

读者可能会认为频率为采样率一半的信号可以由采样值的交替序列，即 0 和 1，精确地表示。如图 20-2 所示，上面一行中采样序列是很清晰的（假定波形的幅度为单位 1）。中间一行的采样序列则可以由图中画出的三个波形中的任意一个得到，下面一行中的相移波形的采样序列与幅度为 0.5 的恒定波形的采样序列没有任何区别。对于频率正好是采样率一半的信号无法进行分析，从而导致采样定理中严格的“小于”条件，如果其相位和幅度是任意的，则无法用采样序列精确表示，下文将会对其进行解释。

哈里·奈奎斯特于 1928 年在贝尔实验室发表了一篇论文指出，为了保证信号采样不会引入混叠，所有信号频率分量必须严格限制在采样频率的一半以内（现在称为奈奎斯特频率）。如果一个信号满足该条件，则称其满足奈奎斯特条件。

1915 年 E. T. Whittaker 取得的一项数学发现，奈奎斯特定理主要应用于信号处理，后期 Claude Shannon (美国) 和 Aleksandr Kotelnikov (俄罗斯) 都对此做出了很多贡献。

这一条件通过在采样前进行模拟滤波实现，即滤掉频率为 $0.5f_s$ 或更高的频率分量。滤波器必须进行某种形式的积分。在图 20-1 所示例子中，并没有使用滤波器，只对波形进行了简单的点采样。不采用滤波器会导致混叠现象。图 20-3 表示一个实际信号的波形；点采样在某一时刻的采样值并不代表该时刻瞬时值。点采样的缺点是，选取了极值点，而极值点无法表示采样点附近的信息。

滤波最基本的方法可能是在每个采样周期中对波形进行平均。有许多不同的可行的积分方案，这些方案可以表示为对时间的加权函数。采样平均使用矩形加权函数表示，如图 20-4 所示，它的值在采样周期内是单位 1，采样周期之外为 0。使用这一加权函数的滤波过程称为矩形（货车）滤波，这是因为这些具有不同幅度的函数序列看上去像一列货运货车。得到了加权值，就得到了该信号的离散值表示，如图 20-5 所示。图中使用矩形函数对输入波形加权，每个输出采样为一个采样间隔上的平均值。

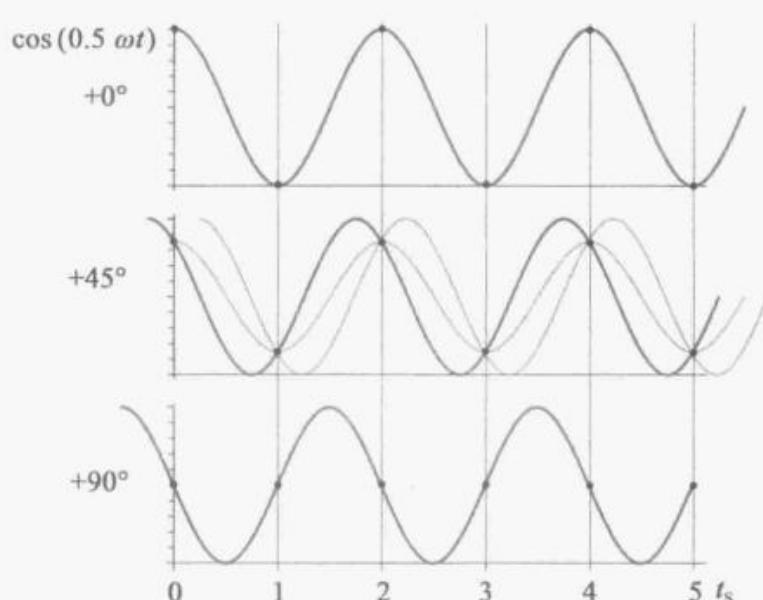


图 20-2 频率正好为 $0.5f_s$ 的余弦波

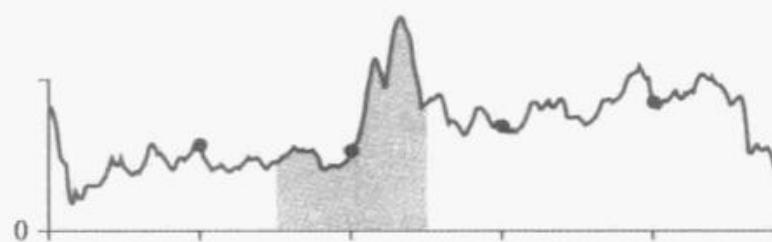


图 20-3 点采样

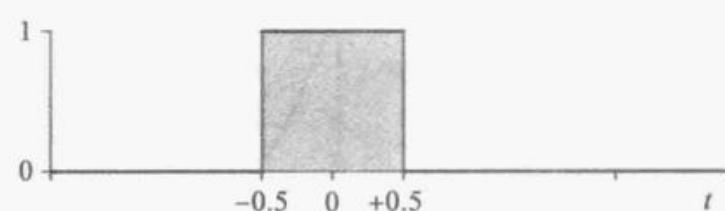


图 20-4 矩形加权函数



图 20-5 矩形加权滤波器

要用矩形函数的幅度来绘制这些数值，则强烈建议用矩形函数作为重建滤波器。图 20-3 所示波形的阴影部分表示矩形滤波。

在采样间隔中使用矩形滤波出现的严重问题在图 20-6 中是显而易见的，上面图像显示的是 $0.75f_s$ 的正弦波，该信号的频率已超过了奈奎斯特频率。阴影区域为一个采样周期间隔的积分。对于频率为 $0.75f_s$ 的正弦波，采样从零相位开始，前两个积分值约为 0.6061，接着的两个积分值约为 0.3939。滤波后的采样序列主值如图 20-6 下方的图所示，其频率为采样频率的 $1/4$ 。采用单个采样周期宽度的矩形加权函数进行滤波是不能够将奈奎斯特频率之上的信号分量充分衰减的，会导致混叠。

图 20-6 给出了另一个混叠的例子：由于使用较差的预采样滤波器，使得采样值序列提供的频率分量并不属于输入信号。正如本例所示，矩形积分并不能保证完全避免严重的混叠。

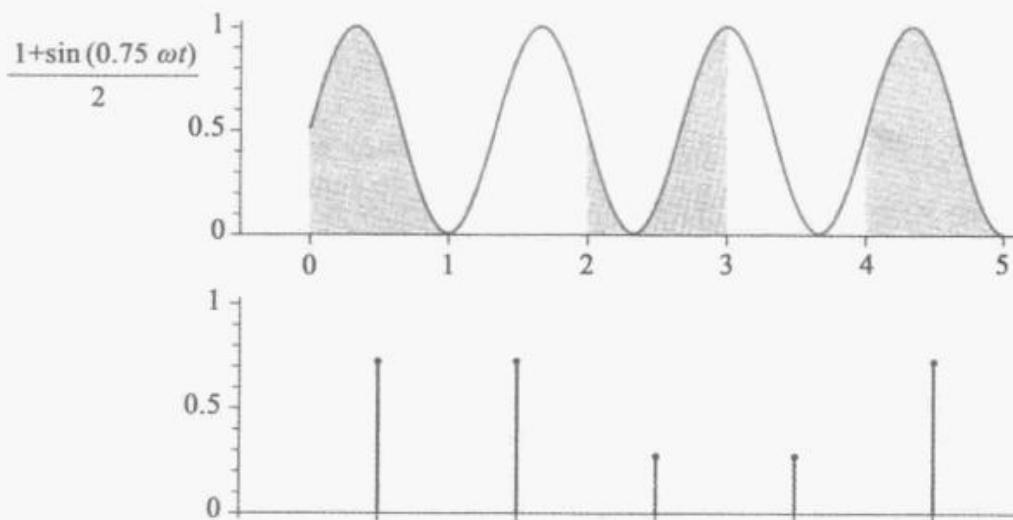


图 20-6 混叠产生的原因是使用矩形滤波

20.3 幅频响应

要得到混叠的一般评价，有必要理解一下频域中的信号。上一节给出了采样之前未能充分滤波而导致采样过程中出现混叠的例子。可以给滤波器提供一个全频段信号，从 0，经低频段，到某个高频段，绘制出想要的滤波器响应，以确定滤波器是否会在某个频率出现不期望的响应。在本书图 9-1 上方的图中，绘出了这样一个频率扫描信号。中间的图显示的是典型低通滤波器（Lowpass Filter, LPF）的频率响应曲线，该滤波器衰减高频信号，其幅频响应如图 9-1 最下面的曲线所示。

严格来说，振幅是一个瞬时测量值，其值可正可负。幅度通常情况下用绝对值，或平方值，或方均根值来表示一段时间的振幅。这两个术语经常互换使用。

幅频响应是输入信号在每个频率的所有相位上的方均根值（RMS）的平均响应。正如上一节所看到的，输入信号的相位对滤波器的响应影响甚大。要确定某个特殊频率的信号响应，可以测试该频率上所有相位的响应。或者，假如滤波器是线性的，可以提供两路信号——处于测试频率的余弦波和具有相同频率的正弦波。滤波器在任意频率上的幅频响应都是余弦波和正弦波响应的向量和的绝对值。

称为变换的解析和数值运算过程可以用来确定频率响应。拉普拉斯变换适用于连续函数，例如在模拟域的信号。傅里叶变换适用于周期采样的信号，或者本身就是周期的信号。用变量对已采样的数据进行运算，即为离散傅里叶变换（DFT）。DFT 变换的简便方案就是快速傅里叶变换（FFT）。 z 变换实质上是傅里叶变换的一般形式。所有的这些变换

显示了用数学方法确定一定范围的频率和相位下的正弦波的系统响应。变换的结果是关于频率的表达式或图形。

20.4 矩形函数的幅频响应

图 20-7 顶部曲线表示的是点采样的加权函数，也是时间的函数（采样间隔内），关于 $t=0$ 对称。矩形函数的傅里叶变换——矩形加权函数的幅频响应——形如 $\sin(x)/x$ 。如图 20-7 底部所示，频率轴单位为 $\omega = 2\pi f_s$ 。实线表示的是在特定频率上，滤波器造成了相位反转。滤波器的响应通常绘成幅频形式；在阻带的相位反转用绝对值（幅度）以虚线的形式表示。式 (20-1) 非常重要，所以用特殊符号 $\text{sinc}(\cdot)$ 表示，由 Phillip M. Woodward 在 1952 年引入，作为下列函数的简写。

$$\text{sinc}\omega = \begin{cases} 1 & (\omega = 0) \\ \frac{\sin(\pi\omega)}{\pi\omega} & (\omega \neq 0) \end{cases} \quad (20-1)$$

$\text{sinc}(\cdot)$ 函数读为 sink，其辐角的单位是弧度每秒 ($\text{rad} \cdot \text{s}^{-1}$)，这里使用的是惯用符号 ω 作为变量。 $(\sin x)/x$ (读为 sine ecks over ecks) 经常用符号 $\text{sinc}(\cdot)$ 表示，而不指出辐角的单位。如果参数是以 Hz 为单位，则该函数可以写为 $\sin(2\pi f)/(2\pi f)$ 。

$\text{sinc}(\cdot)$ 函数与同步无关（同步性）。

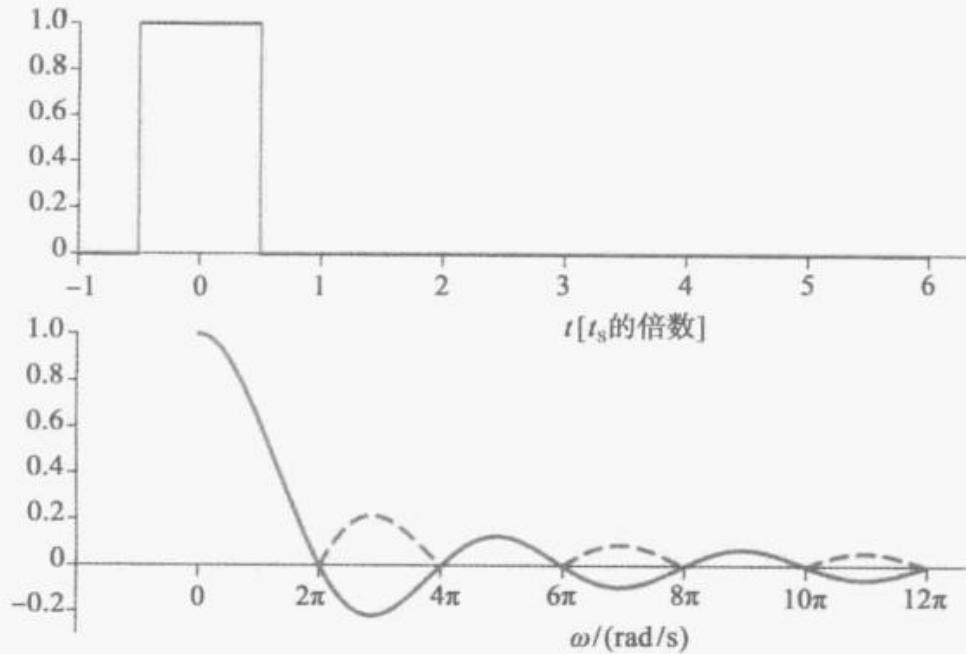


图 20-7 矩形滤波器的频率响应

预采样滤波器在低于采样率一半时，应该有较为一致的响应，并有很好的尖锐度，并且当频率等于和高于采样率一半时需要有较强的频率衰减，以减少混叠。图 20-7 底部曲线显示，矩形加权函数并不能达到这些要求，从 $\text{sinc}(\cdot)$ 函数的图形可以推断出引入混叠的频率段。图 20-6 给出了一个频率为 $0.75f_s$ 的正弦波的例子，从图 20-7 所示的曲线可以读出 $\text{sinc}(\cdot)$ 函数在频率为 1.5π 时的值，可以看出这里将会出现混叠。

考虑如下情况，会对矩形加权函数有个直观的理解：输入信号在矩形窗内为整数倍周期，则响应为零。但如果输入信号在矩形窗内为整数倍周期再加半个周期，则响应将会局部达到最大而引入混叠。

要得到不引入混叠的预采样滤波器，需要让低频信号通过，最大频率不超过采样率

的一半，且将高于该频率的信号滤掉。需要的信号响应为在低于 $0.5f_s$ 时，幅度为常数 1，超过 $0.5f_s$ 时，则幅度下降到 0。需要滤波器函数的频率响应，并非时间响应，这与矩形窗相似。

20.5 $\text{sinc}(\cdot)$ 加权函数

值得注意的是，傅里叶变换具有一种可以成为自身的逆（按一定的比例因子）的数学特性。在图 20-7 中，矩形加权函数的傅里叶变换产生了形状为 $\text{sinc}(\cdot)$ 的频率响应。图 20-8 所示的是对应地显示了一个 $\text{sinc}(\cdot)$ 形加权函数（如图中顶部曲线所示），它产生了一个矩形窗的频率响应（如底部曲线所示），呈矩形。 $\text{sinc}(\cdot)$ 加权函数呈现了预采样滤波器的理想特性。然而，其无线延展性使其在实际中无法实现，加之其负波瓣使其无法应用于光传感器，如照相机、扫描仪、显示器等。许多实际的数字低通滤波器具有类似 $\text{sinc}(\cdot)$ 采样的系数。所以， $\text{sinc}(\cdot)$ 加权函数为一个预采样滤波器， $\text{sinc}(\cdot)$ 是一个理想的时间加权函数。不过， $\text{sinc}(\cdot)$ 函数的使用在理论和实际上都有困难，在实际使用时，只能逼近 $\text{sinc}(\cdot)$ 函数。

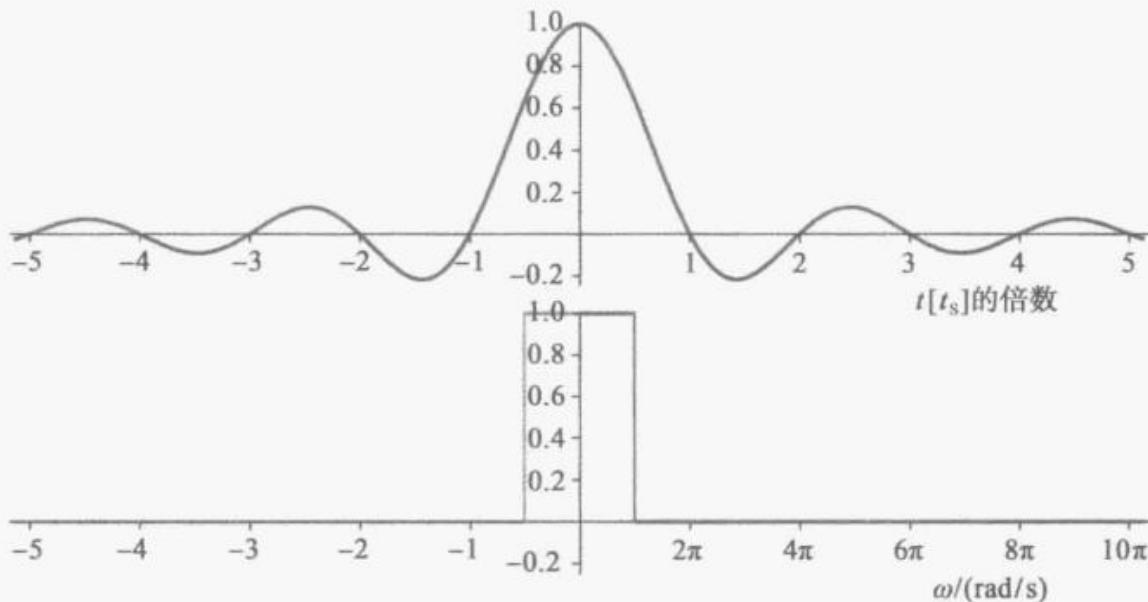


图 20-8 $\text{sinc}(\cdot)$ 加权函数

在模拟视频中接近理想的滤波器有时称为砖墙滤波器，但没有对该术语的精确解释。

模拟滤波器的响应是在正实轴上的频率函数。在模拟信号理论中，频率没有上界，但是在数字滤波器中，测试频率 f_T 的响应，与测试频率偏移整数倍的采样频率所得到的响应是完全相同的：频率轴“涵盖”多倍的采样率。采样定理规定在采样频率约一半处进行“折叠”。那些等于或大于奈奎斯特频率的信号分量将无法准确地表达出来。

视频上使用的时间加权函数通常是对称的，然而，其通常用双边形式的图形画出来。适用于实信号的滤波器的频率响应近似关于零点对称。习惯上，将频率响应画为从零频（“直流”）开始的单边曲线，有时将频率响应考虑为或画成双边形式是很有用的。

20.6 点采样的频率响应

傅里叶变换提供了一种分析工具来检测频率响应：可以重新检测点采样。对波形进行瞬时采样在数学上等同于使用采样时刻为单位 1，其他时刻均为零的加权函数——即加权函数是一个脉冲。冲激函数的傅里叶变换在整个频域上都是一个常数 1。一串等间隔的脉冲为脉冲序列，其傅里叶变换为单位 1。采样过程可以表示成与一个脉冲序列相

乘。一个未经滤波的信号经脉冲序列采样后会受到来自于所有的输入信号频率的同等的混叠影响。

20.7 傅里叶变换对

高斯函数如图 20-9 所示。这里给出的高斯函数是单侧的，采用统计的缩放比例，该函数（以均值和方差为参数）称为正态函数。它的积分是误差函数 $\text{erf}(x)$ 。级联滤波器的频率响应也是高斯函数。高斯函数是傅里叶变换的恒等函数，它具有其变换仍为其自身的独特特性（以一定的缩放因子）。高斯函数在时域和频域上都具有适度延伸的特性，其取值范围为无限长，但是离开原点很小单位取值就变得小到可以忽略不计。高斯函数处在时域能量分布和频域能量分布的平衡点上。

具有较短持续时间的函数可以变换成具有较宽分布的频率分量。反过来，频率较为紧密的函数可以变换成持续时间较长的时域函数，如图 20-10 所示。图 20-10 所示的三个波形的时域图形显示在左侧，相应的变换显示在右侧。频谱宽度与时间范围的长度成反比，不只是这里显示的高斯分布，所有的波形都是如此。

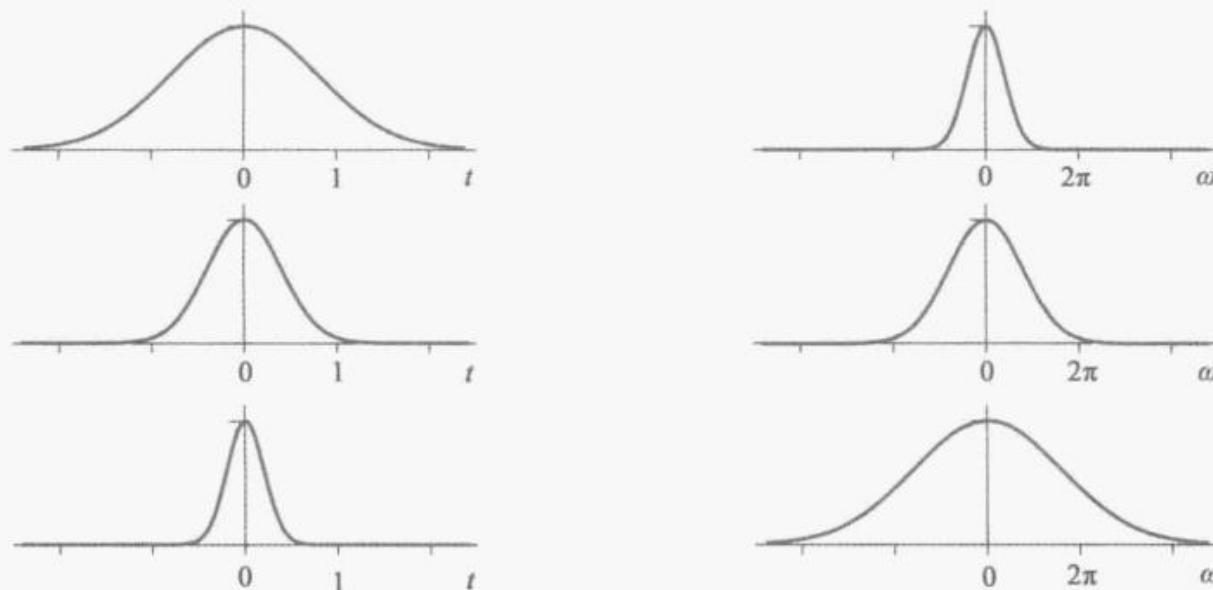


图 20-10 三个波形的时域图形

图 20-11 显示了几个不同函数的傅里叶变换对，左边一列是一组波形，波形中间那行是高斯分布函数，右边一列是相应的频谱。

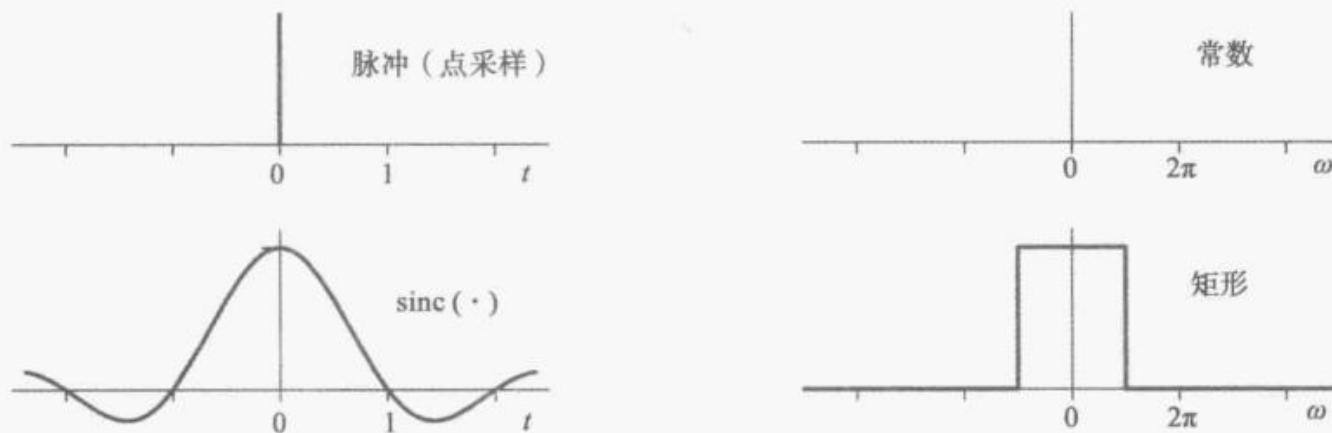


图 20-11 傅里叶变换

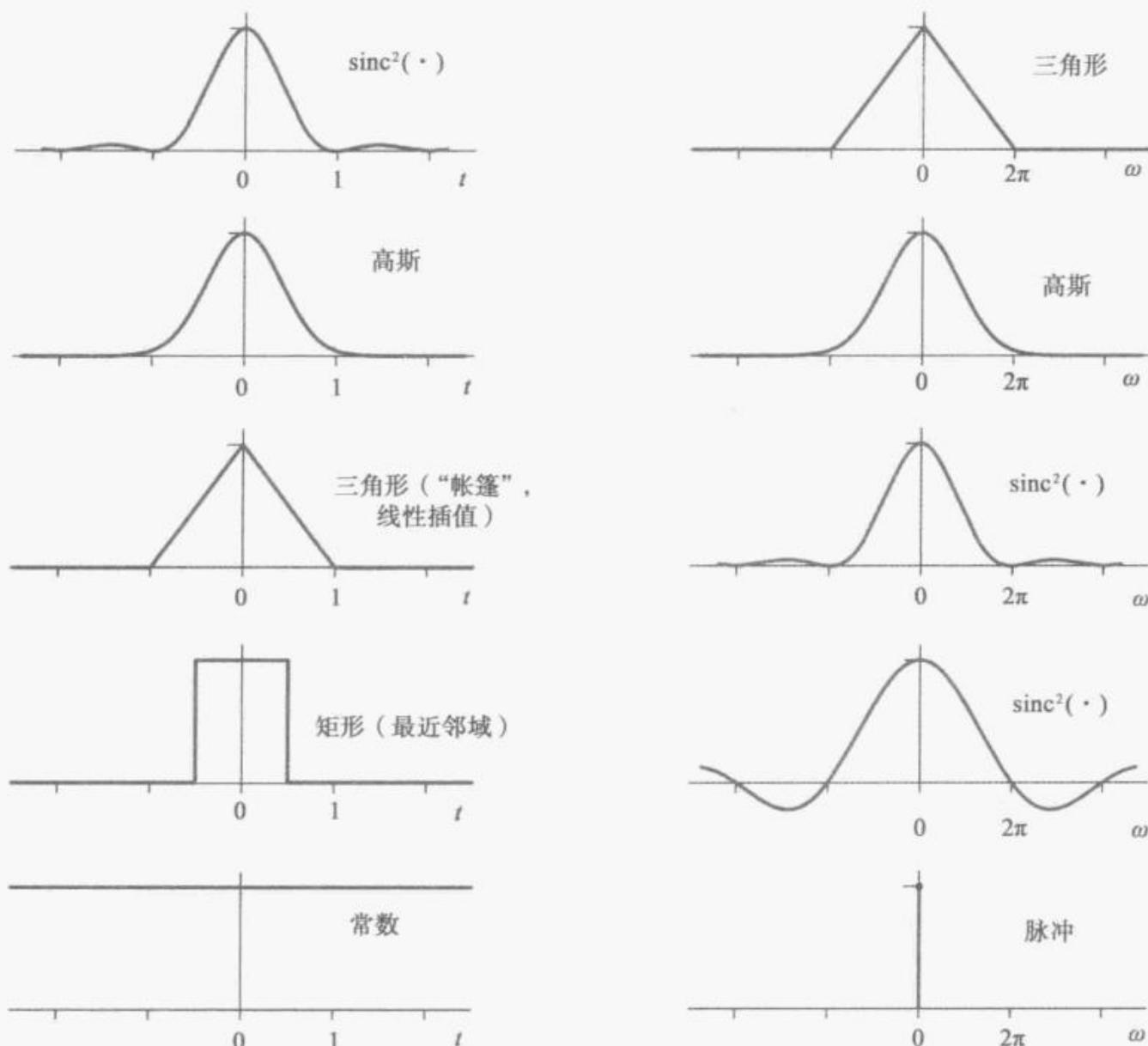


图 20-11 (续)

20.8 模拟滤波器

模拟滤波是数字化前的必要步骤，是要将信号在不发生混叠的情况下引入数字域。前面已经介绍了使用不同加权函数积分从而实现滤波的方法；抗混叠滤波器使用模拟电路实现积分运算。

模拟滤波器利用存储磁场的电感器（线圈）的电感（L）特性，或者存储电荷的电容器的电容（C）特性来实现积分运算。在低性能的滤波器中，也使用了电阻（R），普通的模拟滤波器若具有脉冲响应，则具有无限的时域范围。

模拟滤波器最好由专业人员来设计。

20.9 数字滤波器

信号数字化后就可以直接在数字域进行滤波。这样的滤波器的设计和实现——硬件、固件或软件——则是数字信号处理（DSP）的领域了。本章所介绍的滤波器都是通过计算采样点的加权和来实现的。

也许最简单的数字滤波器就是将相邻采样点求和的滤波器；在这种情况下，权重是（1, 1）。图 20-12 显示的便是这样的（1, 1）滤波器的频率响应。该滤波器对极低的频率会产生极小的衰减；当信号频率接近采样率的一半时，响应服从余弦曲线减小到 0。这是

一种非常简单、廉价的低通滤波器 (LPF)。

对邻域采样点加以平均是移动平均滤波器 (MA) 最简单的形式。

图 20-12 所示的灰色曲线表示了从 $0.5f_s$ 到采样频率这一段的滤波器响应。在数字滤波器中，这段频率是无法与 0 到 $0.5f_s$ 段的频率区分开的。零频 (直流 (DC)) 时滤波器的增益为 2，即其系数之和。通常，这样的滤波器系数会进行归一化处理，这样使滤波器的总直流增益为 1。这样归一化系数为 $(1/2, 1/2)$ ，然而，称其为 $(1/2, 1/2)$ 滤波器太不方便，通常称为 $(1, 1)$ 滤波器。

数字滤波器可以用软件、固件或硬件实现。本书在每个图的右上方，都画出了硬件设计者熟悉的框图。

每个标记 R 的方块代表一个寄存器；一系列这些元件形成一个移位寄存器。

简单的高通滤波器 (HPF) 是将前一个采样点减去当前采样点构成的：该滤波器的权重系数为 $(1, -1)$ 。该滤波器的响应如图 20-13 所示。一般在这种情况下，高通滤波器是由未经滤波器的信号减去低通滤波后的信号得到的。未滤波的信号可以认为是权重系数为 $(1, 0)$ 的双抽头滤波器，减去权重系数为 $(1/2, 1/2)$ 的低通滤波器得到权重系数为 $(1/2, -1/2)$ 的高通滤波器。

信号经低通滤波后减去信号自身相当于对低通滤波器中所有系数取反，中心系数除外，它是用单位 1 去减。

图 20-14 显示的滤波器响应是当前采样值与其前面第二个采样值相加，并忽略中间采样值得到的响应。在这种情况下，形成了一个简单的权重系数为 $(1, 0, 1)$ 、在 $0.25f_s$ 频率处的带阻滤波器 (Bandreject Filter, BRF)，也称为陷波滤波器。这里，该滤波器在 $1/4$ 采样频率时的响应为零。这种缩放滤波器通过直流信号时没有衰减。它会使图像数据变得一团糟——如果一个尖桩篱笆要通过这个滤波器，而尖桩正好位于 $0.25f_s$ 频率处，那么数据包会整体求平均而消失！对图像数据采用这样的滤波器是个坏主意，但是这种滤波器（类似的滤波器）对于信号处理是非常有用的。

当带通（带阻）滤波器的通带（阻带）涵盖了频率（倍频）范围的 2 : 1 或更少时，这种带通滤波器可认为是窄带滤波器。

图 20-15 显示了一个滤波器的响应是前面第二个采样值减去当前采样值，并忽

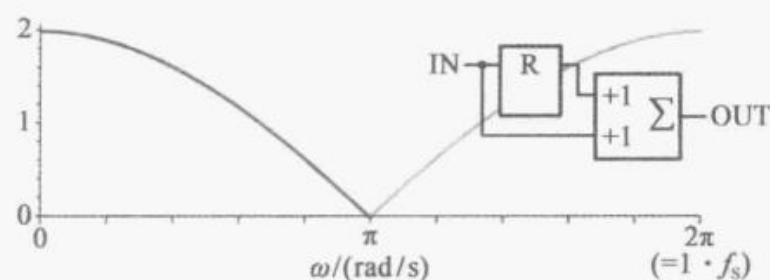


图 20-12 (1, 1) 滤波器

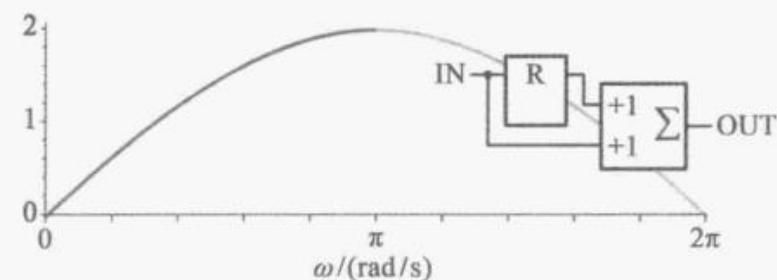


图 20-13 (1, -1) FIR 滤波器

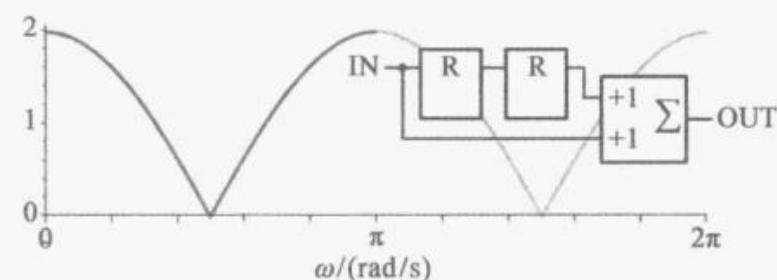


图 20-14 (1, 0, 1) FIR 滤波器

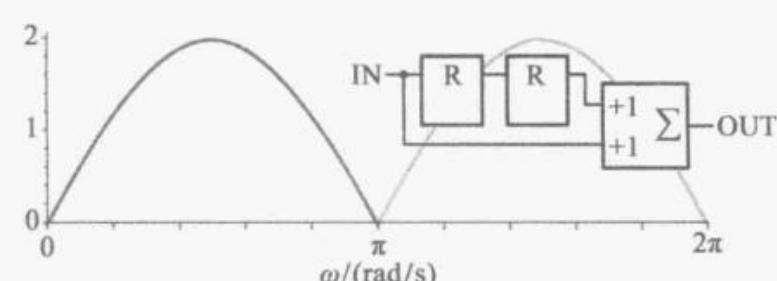


图 20-15 (1, 0, -1) FIR 滤波器

略掉中间值而得到响应。其权重系数为 $(1, 0, -1)$ 。这形成了一个简单的中心在 $0.25f_s$ 的带通滤波器 (Bandpass Filter, BPF)。其权重系数之和为 0——这种滤波器不允许直流通过。该例中的带通滤波器是 $(1, 0, 1)$ 滤波器的互补形式。

图 20-16 显示了一个 5 抽头有限冲激响应 (FIR) 滤波器的框图，包括 4 个寄存器和一个加法器；该滤波器对连续 5 个采样值求和。在归一化之前，权重系数为 $(1, 1, 1, 1, 1)$ 。图 20-17 所示的浅灰色曲线，可得到一个低通滤波器。其频率响应有两个零点，任何频率在 $0.2f_s$ 和 $0.4f_s$ 的输入信号响应都为零，在频率为 $3/10$ 采样频率处阻带的衰减仅有 -12dB 。

如果图 20-16 所示的滤波器具有很多抽头，则需要很多加法器。其算术运算可以通过累加器得到输入样本的运行总和来简化。另外两个累加器计算移位寄存器中的输出运行总和，减法器得到两个运行总和的差值。这种结构称为梳状级联积分器 (CIC)。

在数字滤波器的设计中，对频率响应的控制体现在抽头权重的选择上。图 20-18 显示了具有从高斯波形中选择的分数系数的滤波器框图，其系数为 $(13, 56, 118, 56, 13)$ ，比例系数为 $1/256$ 。系数近似为高斯分布，所以频率响应也近似为高斯分布。乘法器可以通过查表实现。图 20-17 所示洋红色曲线显示这一组抽头权重可得到具有高斯频率响应的低通滤波器。使用负的系数，低频段的响应可以扩大到高频段而不会降低性能。图 20-17 所示浅灰色曲线显示了权重系数为 $(-32/256, 72/256, 176/256, 72/256, -32/256)$ 的滤波器响应。该滤波器在高频段（约 -18dB ）上作为高斯波形会有相同的衰减，但在 -6dB 频段会有约两倍的衰减。

在最后一个例子中，负系数可能导致采样值大于单位 1（在该例中，大于单位 1 的输出采样值会出现在 $\omega=0.3\pi$ ，采样率的 $1/6$ 处）。如果极值被省略，伪值就会出现。为了避免出现伪值，信号编码范围要包括适合的上余量和下余量。

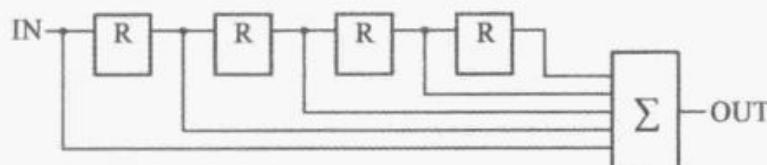


图 20-16 一个非常简单的 5 抽头 FIR 滤波器

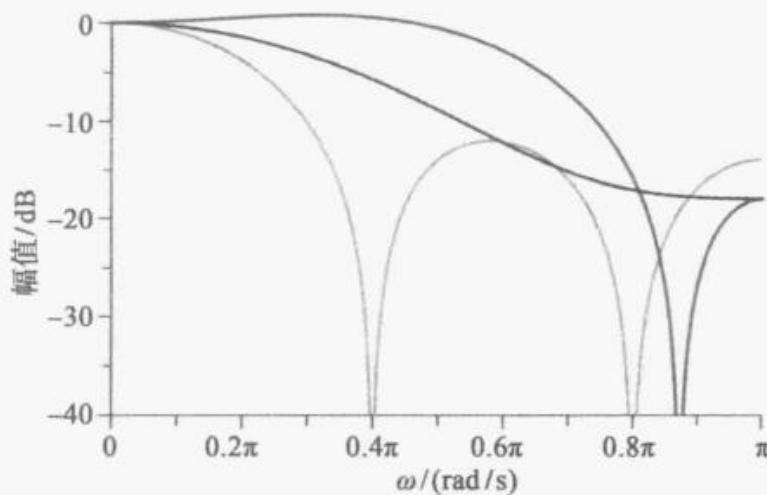


图 20-17 5 抽头 FIR 滤波器的响应

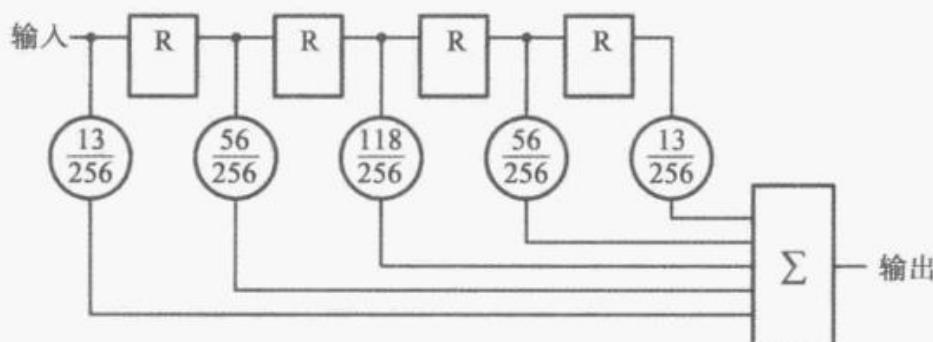


图 20-18 包含乘法器的 5 抽头滤波器

FIR 滤波器的操作相当于一组输入采样点与一组滤波器系数（权重）相乘，得到相应的一组乘积和。权重可以通过乘法器或者查表法得到。对一个完整的输入样本集来说，该

操作称为卷积。通常，卷积可以认为是每次进行一次乘法，一个 n 抽头 FIR 滤波器可以使用单个运行在 n 倍采样率下的乘法器-累加器（MAC）结构实现。用 n 个乘法器结构直接实现，或者单个 MAC 的复用实现，输入样本的接收和输出样本的输出均按时序进行：每个系数需要经过滤波器 n 次。然而，卷积是关于输入样本和系数对称的，即同一组结果可以通过每次向 MAC 输入一个滤波器系数，且对每个输入样本累加部分和来实现。FIR 有许多可能的实现结构。

如欲了解实现结构的细节，请参阅本章结尾处引用的 Lyons 和 Rorabaugh 所著的书。

图 20-19 显示了一个具有 8 抽头、权重系数为 $(1, 0, 0, \dots, 0, 1)$ 的 FIR 滤波器的框图，包括一些延时单元和一个加法器。该滤波器的频率响应如图 20-20 所示。一个正好为周期整数倍的信号经过滤波器时响应达到峰值；当为整数倍外加半个周期时，滤波器响应为零。所有的峰值均具有相同的幅度：当正好有 1, 2, 3\dots 或 n 样本位于加权窗内时，响应是相同的。这种滤波器的幅频响应的形状如同梳子，所以这种滤波器称为梳状滤波器。该滤波器在零频上有归一化的响应：允许直流通过。权重系数为 $[1/2, 0, 0, \dots, 0, -1/2]$ 的滤波器会阻隔直流。

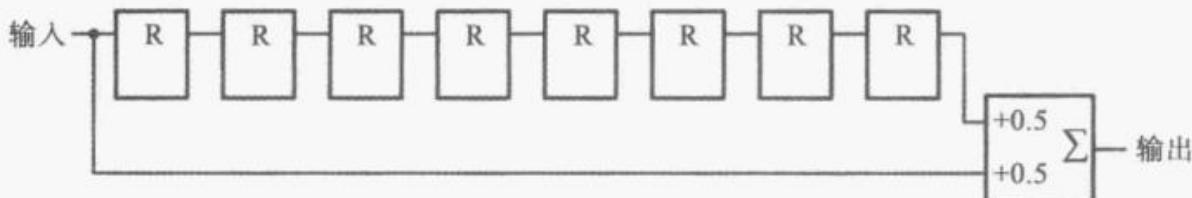


图 20-19 一个简单的梳状滤波器

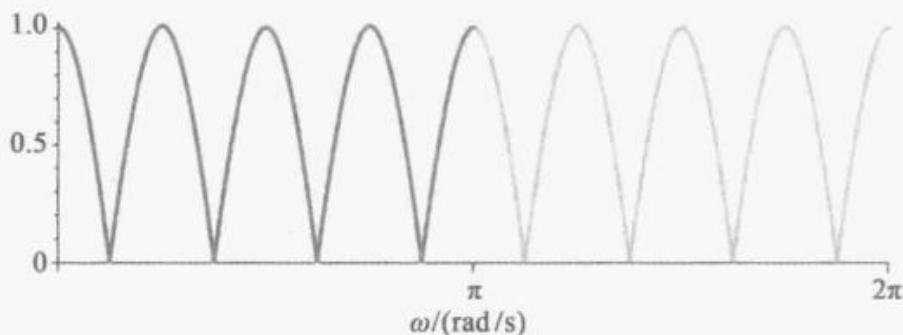


图 20-20 简单的梳状滤波器的响应

20.10 冲激响应

我已经将滤波解释为沿着时间轴的加权积分，并引入了时间加权函数来表示权重。我认为对于一个数字技术专家来说，对滤波在时域上操作的解释会比由频域入手的传统解释更为直观。但我引入的术语——时间加权函数是不标准的，现在我必须介绍常见但不直观的术语——冲激响应。

模拟的冲激信号具有无限长的持续时间、无限大的幅度，且积分为 1（模拟冲激信号概念上等同于数学上的狄拉克或者克拉内罗 δ ）。数字冲激信号是位于一串零值中的具有单位幅度的单个样本；数字滤波器的冲激响应是除单个单位幅度样本之外样本值均为零的输入得到的响应。

20.11 有限冲激响应滤波器

到目前为止本章所介绍的滤波器中，只有几个的系数是非零的。当一个数字脉冲通过这类滤波器时，其结果是简单的依次输出的权重系数。冲激响应的持续时间是有限

的，前面所举的例子就是有限冲激响应。它们是 FIR 滤波器，在这种滤波器中，频率响应等于滤波器的系数集合。20.9 节介绍的数字滤波器直接在时间上进行加权。这些滤波器的冲激响应，进行归一化后，分别是 $(1/2, 1/2)$, $(1/2, -1/2)$, $(1/2, 0, 1/2)$ 和 $(1/2, 0, -1/2)$ 。

图 20-18 所示的特定的一组加权值近似为采样高斯波形，所以该滤波器的频率响应也近似为高斯波形。该滤波器的操作可以用公式表示为：

$$g(j) = \frac{13}{256}s_{j-2} + \frac{56}{256}s_{j-1} + \frac{118}{256}s_j + \frac{56}{256}s_{j+1} + \frac{13}{256}s_{j+2} \quad (20-2)$$

我所描述的冲激响应在某一时间点上是对称的。你可能认为 $t=0$ 应该是时间起点，但是通常为了方便平移时间轴， $t=0$ 点会对应于滤波器冲激响应的中心点。FIR（或者非递归）滤波器只有有限几个系数是非零的。当冲激信号超出这个区间时，响应为零。多数用于视频的数字滤波器都是 FIR 滤波器，大多冲激响应都是关于 $t=0$ 对称或者反对称的。

在式 (20-2) 中， g 是一个序列（其序号标示在方括号中），而不是一个函数（其参数在圆括号内）， s_j 是第 j 个样本。

偶对称为： $f(x) = f(-x)$

奇对称为： $f(x) = -f(-x)$

可以将 FIR 滤波器视为固定结构，数据沿着滤波器结构移动。或者，可以认为数据是固定的，滤波器在数据上滑动。这两种观点是等价的。

20.12 滤波器的实际可实现性

为了应用在实际中，数字滤波器必须是实际可实现的：在有限持续时间内有一个时间加权函数（冲激响应）是实际需要。FIR 滤波器需要存储若干个输入样本，每个采样周期内需要进行多次乘法操作。所存储的输入样本的数目称为滤波器的阶数，或者称为滤波器抽头的个数。如果一个特定的滤波器系数固定，那么它的乘法运算可以通过查表来进行。一个简单的技巧是利用冲激响应的对称性来消除一半的乘法运算，这通常是有利的！

当一个时间加权函数经过某个点被截断后，它的变换——频率响应特性都会受到影响。滤波器设计的技术和技巧包括选择滤波器的阶数，也就是超出加权函数的部分会被强制归零。这部分需要离中心抽头足够远，如此，滤波器高频响应就会足够小从而可以忽略。

这里用了“截断”这个词来表示加权函数超过固定抽头数后的强制归零。加权函数中的非零系数可能包含一些已量化为位数据的理论值。系数量化可以通过舍入或截断来完成。注意区分冲激响应的截断和系数的截断。

信号处理包括具有负值的冲激响应的使用，负系数在数字信号处理中是常见的。但是图像采样和图像显示涉及感光和生成光线，没有负功率，所以负的权重系数总是无法实现的。学习了傅里叶变换对后，就会发现，如果不能使用负的权重系数，则调整滤波器频率响应的能力会严重受限。

冲激响应在 FIR 数字滤波器的设计中一般是非常明显的。在模拟域上实现矩形滤波器是可行的，模拟滤波器很少直接使用时间加权系数，它的实现一般和冲激响应有着不明显的关系。模拟滤波器最好是从拉普拉斯变换的角度来诠释，而不是用傅里叶变换。模拟滤波器的冲激响应很少直接在设计过程中加以考虑。尽管主要概念和实现过程不同，模拟滤波器与 FIR 滤波器，以及 IIR 滤波器（将会讲到）的特性都由其频率响应确定。

20.13 相位响应（群延迟）

到现在为止，本章已经介绍了滤波器的幅频响应。相频响应——也称为相位响应——也很重要。考虑一个对称的 FIR 滤波器有 15 个抽头，不管输入信号怎样，输入响应都会有 8 个采样周期的延迟，与滤波器冲激响应的中心样本相一致。FIR 滤波器的时间延迟是常数，与频率无关。

考虑一个频率为 1MHz 的正弦波，另一个正弦波频率也为 1MHz，但延迟 125ns。情况如图 20-21 所示。125ns 的延迟在频率为 1MHz 时表现为 45° 的相位平移。但是，如果时间延迟保持不变而频率加倍的话，则相位偏移会加倍，达到 90°。当时间延迟不变时，相位偏移的增长与频率增长成正比（线性）关系。因为在这种情况下，相位延迟与频率是成正比的，故也称为线性相位。恒群延迟（constant group delay）是相位延迟和频率密切相关的一种情况，其中延迟的一阶导数是常数，但也可能存在固定的时间延迟。所有的 FIR 滤波器都会有恒群延迟，但只有对称 FIR 滤波器具有严格的线性相位。

这是许多滤波器的特性——比如说无限冲激响应（IIR）滤波器，后面将会讲到——但是对于不同频率函数，延迟略有不同。一个图像信号含有由不同尺度的场景元素产生的多种频率。如果复制对象的水平位移是取决于频率的话，讨厌的伪影就会产生。对称的 FIR 滤波器在其通带内表现出线性相位，避免出现伪影。在图像处理和视频中，FIR 滤波器的性能会比其他滤波器要好很多：在视频系统中线性相位是一个非常理想的特性。

20.14 无限冲激响应滤波器

目前为止所介绍的数字滤波器都是 FIR 类的。第二类数字滤波器的特性是具有潜在的无限冲激响应（IIR）。IIR（或递归）滤波器计算输入样本加权和——这和 FIR 滤波器情况相同——不同的是将其与先前输出样本加权和相加。

信号处理工程中的 IIR 滤波器，在金融和统计领域称为自回归移动平均（Autoregressive Moving Average, ARMA）。

如图 20-22 所示的为一个简单的 IIR 滤波器：输入样本的权重为 1/4，先前输出样本的权重为 3/4。这些加权值求和得到滤波结果。然后将滤波结果反馈到输入端进行后续样本的计算。输入脉冲进入后，冲击响应迅速出现，并在多个采样周期后渐渐消失。这是一个简单的单抽头低通滤波器，其时域响应与模拟 RC 低通滤波器相似。将先前存储的滤波结果减去输入样本就得到一个高通滤波器（每一个 IIR 滤波器都是非线性相位的）。

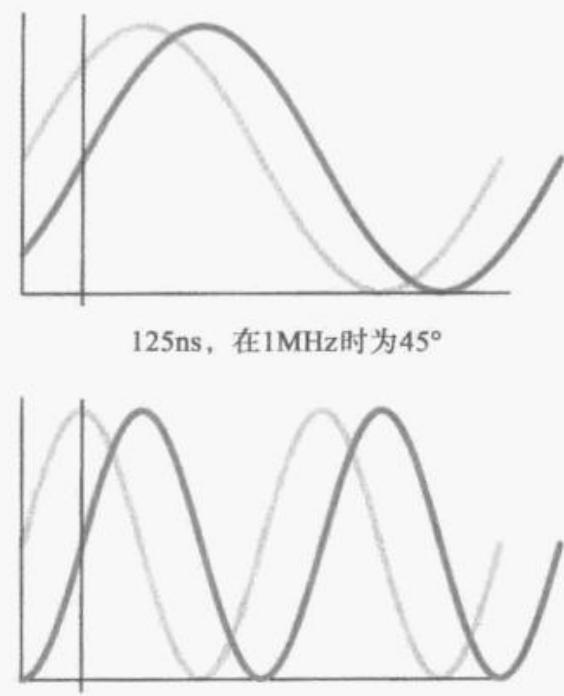


图 20-21 线性相位

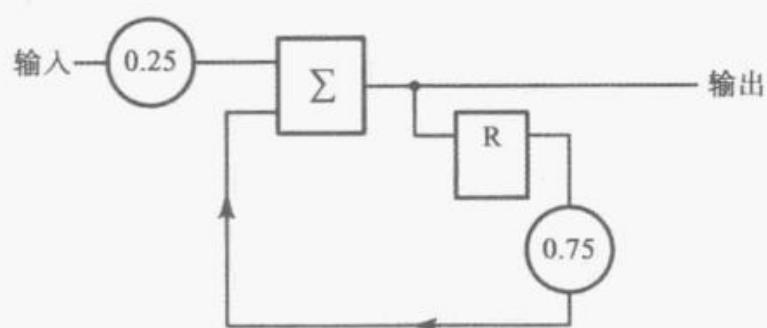


图 20-22 IIR（“递归”）滤波器

在只有一个抽头的 IIR 滤波器中，设计者调整频率响应的能力会严重受限。IIR 滤波器可以通过存储多个先前滤波结果来扩展，并将先前滤波结果的一部分与当前输入样本的一部分相加（或相减）。在多抽头 IIR 滤波器中，要对频率响应实现精细的控制可以通过使用少数抽头实现。在 IIR 滤波器中只需三四个抽头就可以完成的频率响应，在 FIR 滤波器中可能需要 20 个抽头。

然而有一个问题：IIR 滤波器的衰减和延迟都取决于频率。用上一节提到的术语，IIR 滤波器是非线性相位的。通常情况下，低频信号会比高频信号延迟更大。正如前面所解释的，延迟随频率变化而变化的函数关系可能是视频中一个严重的问题。

尽管复杂的 IIR 滤波器可以将相位误差设计成任意小，但其不能具有精确的线性相位。因为 IIR 滤波器通常具有较差的相位响应，所以不常用于视频（一个值得注意的例外是，用在减小时域噪声消除时使用基于场和基于帧的 IIR 滤波器，其中延时器件中会包括一场或一帧的存储）。

滤波器中不需要的相位补偿称为均衡。这与同步中均衡冲激无关。

由于 IIR 滤波器的输出结果依赖于先前的输出结果，故 IIR 滤波器必然是递归的。然而，有些递归滤波器具有有限长的冲激响应，所以递归滤波器不一定都具有无限长的冲激响应。

术语非递归和递归最好用于描述滤波器的实现结构。

20.15 低通滤波器

低通滤波器允许低频信号原样通过，但是衰减高频信号。图 20-23 描述了一个低通滤波器，响应具有通带，其中滤波器响应接近归一化的 1；有过渡带，即滤波器的响应为中间值；有阻带，即滤波器的响应近乎为零。对于低通滤波器，转角频率 ω_c ，有时也称为带宽或截止频率，是滤波器的幅频响应由其参考频率（通常是零或是直流）的幅度下降 3dB 时的频率。换句话说，在转角频率上，滤波器的响应下降到其直流响应的 0.707，输出功率为输入功率的一半。

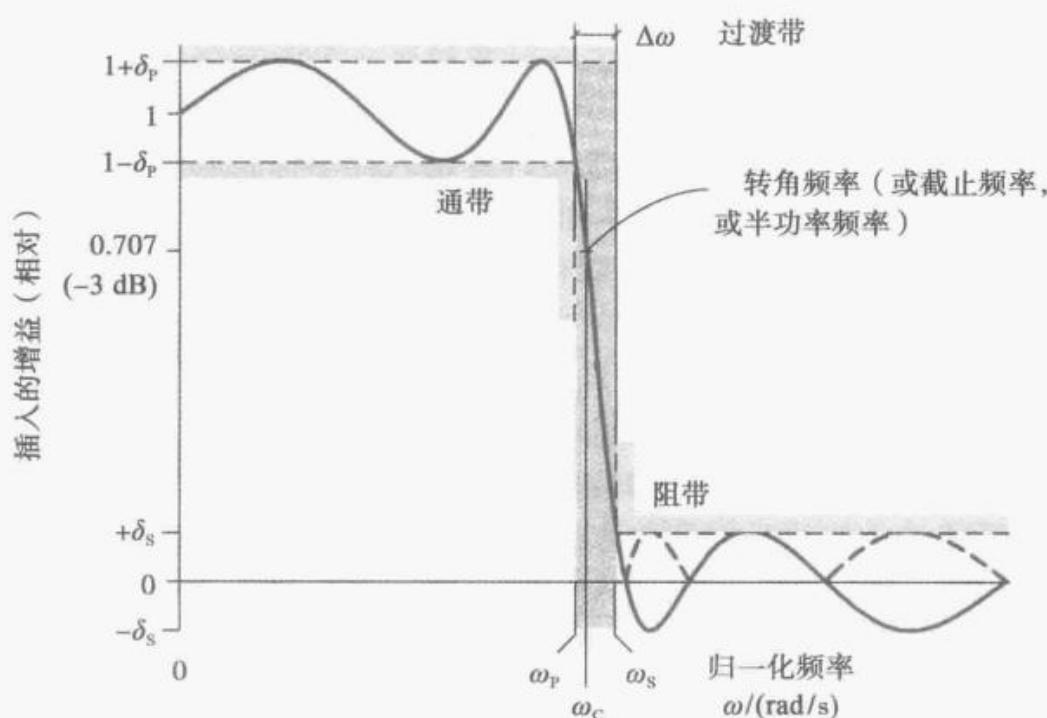


图 20-23 低通滤波器的特性

其中，用符号 ω 表示频率，其单位是弧度每秒 (rad/s)。数字滤波器频率随着采样频

率变化而变化，故使用 ω 较为方便，因为采样频率始终为 $\omega = 2\pi$ ，半采样频率（奈奎斯特频率）总为 $\omega = \pi$ 。

通带用通带边沿频率 ω_p 和通带纹波 δ_p （有时也用 δ_1 表示）来描述。在通带上，响应在 δ_p 内为 1，误差在 1% 左右。阻带用阻带边沿频率 ω_s 和阻带纹波 δ_s （有时也用 δ_2 表示）来描述。阻带上，响应在 δ_s 内为零，误差在 1% 左右。过渡带位于 ω_p 和 ω_s 之间，其宽度为 $\Delta\omega = \omega_s - \omega_p$ 。

低通滤波器的复杂程度主要取决于其归一化的过渡带宽的宽度（或过渡比） $\Delta\omega/2\pi$ 。过渡带宽越窄，滤波器越复杂。此外，无论是通带还是阻带，纹波越小，滤波器复杂度越高。FIR 滤波器抽头的数量可以根据 Bellanger 公式进行如下估算：

$$N_e \approx \frac{2\pi}{\Delta\omega} \times \frac{2}{3} \times \lg\left(\frac{1}{10\delta_p\delta_s}\right) \quad (20-3)$$

在模拟滤波器设计中，频率响应通常在双对数坐标中绘制，频率轴的单位是 Hz 的对数，幅度响应的单位是分贝（dB）。在数字滤波器设计中，频率通常从零到采样率的一半按线性方式绘制。数字滤波器的通带和阻带响应通常都是用对数方式绘制的；通带响应通常被放大，用于强调对于单位 1 的微小偏离。

在 BT. 601 中用于专业数字视频预采样滤波器的标准模板如图 20-24 所示。顶部的模板显示了频率响应，详细描述了通带（在顶部）和阻带细节（在中部）。底部模板显示的是群延迟说明。符合该模板的实际低通滤波器如图 20-25 所示。图 20-25 所示的为一个半带滤波器，使用的采样频率为 27MHz、转角频率为 0.25fs。实用级滤波器可能会有比这两个数量级的纹波。图 20-25 所示的为一个实际滤波器的频率响应，其转角频率是采样频率 27MHz 的 1/4。图 20-4 所示的横坐标（频率）是线性的，纵坐标（响应）是对数

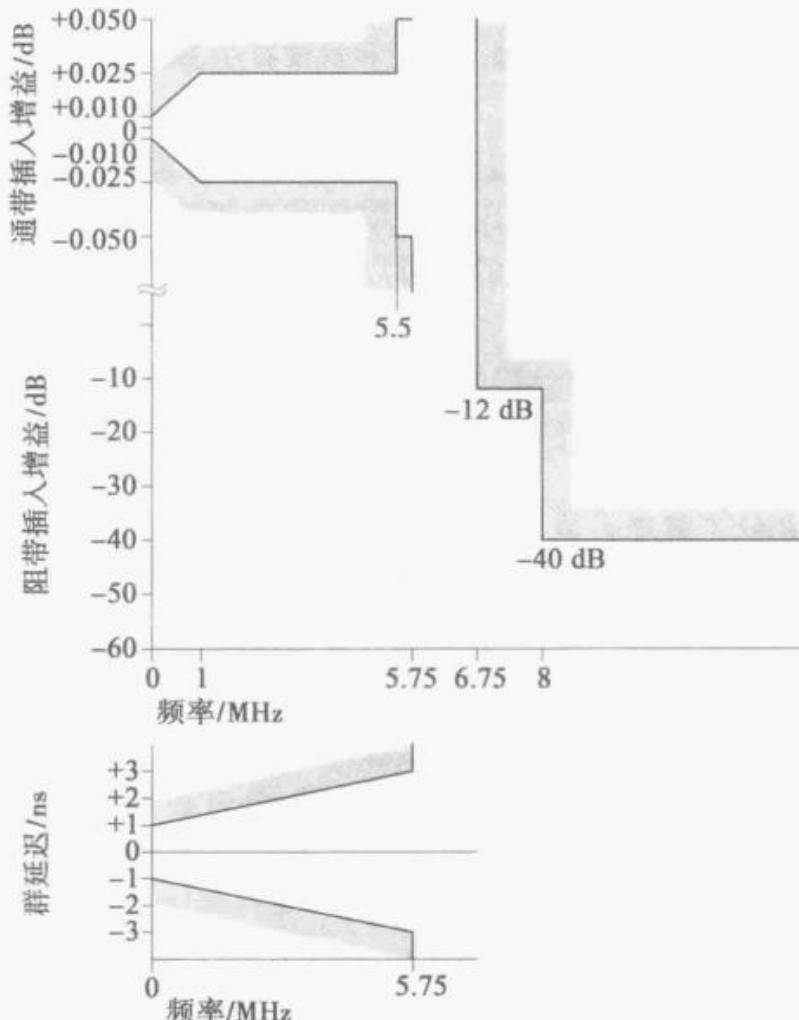


图 20-24 BT. 601 滤波器模板

的。顶部的图显示的是通带具有总体增益 1 和约 ± 0.02 dB 的一致性（纹波）：通带中，增益范围在 0.0997 和 1.003 之间。底部图像显示的是阻带约有 -60 dB 的衰减阻隔：滤波器在该段频率上约有 0.001 的增益。这些 GF9102A 半带滤波器的数据是由加拿大数字芯片公司友情提供的。

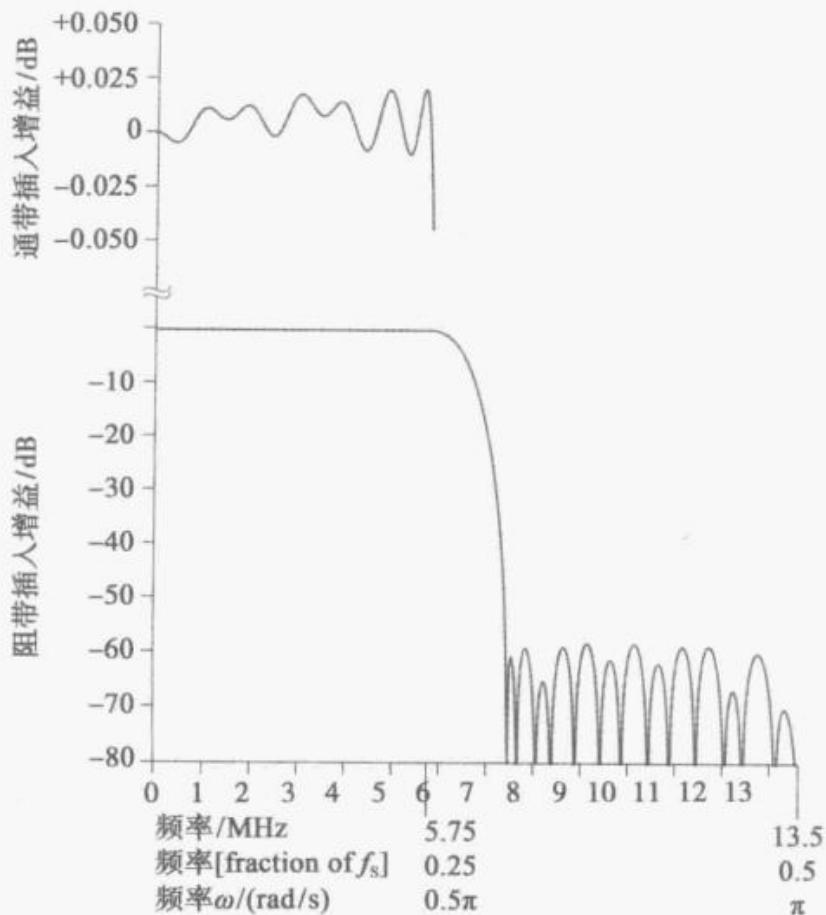


图 20-25 半带滤波器

20.16 数字滤波器设计

设计数字滤波器简单的方法是，使用合适数量的理想冲击响应的采样点的系数。系数超过某个点——滤波器的阶数——后面的就可以忽略掉。式 (20-4) 实现了近似高斯采样的 9 抽头滤波器：

$$g(j) = \frac{1s_{j-4} + 9s_{j-3} + 43s_{j-2} + 110s_{j-1} + 150s_j + 110s_{j+1} + 43s_{j+2} + 9s_{j+3} + 1s_{j+4}}{476} \quad (20-4)$$

本书对上升时间进行了解释，在阶跃输入的响应中，高斯滤波器在转角频率处有一个非常接近于 $1/3$ 周期的上升时间。

系数的忽略会造成频率响应与理想情况有一定的偏离。如果忽略的系数比 0 大很多，实际的频率响应可能会与理想情况相差甚大。

另一种设计数字滤波器的方法是从理想低通滤波器开始。它的无限延伸可以通过简单截断其权重系数来处理，即迫使权重系数归零，除了一定的时间间隔内之外，比如说， 0 ± 4 个采样周期。这对频率响应会有不好的影响，不过，频率响应会在过渡带附近出现过冲和下冲。

截断的 $\text{sinc}(\cdot)$ 函数的较差的谱状态可以通过加权函数使得峰值出现在滤波器中心，使间隔边界平缓归零来改善。这过程称为使用窗函数。使用窗函数设计滤波器时，先使用 $\text{sinc}(\cdot)$ 函数在时间轴进行缩放，来选择转角频率和合适的抽头数。每个抽头的权重系数

用 $\text{sinc}(\cdot)$ 函数值与相应的窗函数的值相乘来得到。 $\text{sinc}(\cdot)$ 函数可以通过乘以一个矩形窗函数进行截断。也许最简单而重要的窗就是三角形窗，也称为巴特利特窗。汉宁窗（经常误称为“汉明窗”）是一个单周期的升余弦函数。像汉宁窗这样的窗函数是由转角频率和抽头数固定的，无法控制过渡带的宽度。

可以用“加权”这个词，但是 $\text{sinc}(\cdot)$ 函数本身就是一个加权函数，所以这里选择了一个不同的词：加窗。

凯塞窗拥有一个单独的参数来控制它的宽度。对于给定的滤波器阶数，如果过渡频带变窄，则阻带衰减就会减小。凯塞窗的参数允许由设计人员来决定如何进行权衡。

加窗的 $\text{sinc}(\cdot)$ 函数要比截断的 $\text{sinc}(\cdot)$ 函数具有更好的性能，而且窗函数很容易设计，所以没有理由在使用 $\text{sinc}(\cdot)$ 函数时不加窗。然而在多数工程应用中，滤波器的性能在频域中得到最好的表现，加窗 $\text{sinc}(\cdot)$ 函数在频域上的性能并未达到最好： n 抽头的加窗 $\text{sinc}(\cdot)$ 滤波器的性能远不如经过合理优化后的 n 抽头 $\text{sinc}(\cdot)$ 滤波器。

少数几种封闭式方法是设计最优滤波器的方法。设计一个高性能滤波器的方法通常包括逐步逼近的方法，通过在时域频域间反复训练设计参数进行优化。由 J. H. McLellan, T. W. Parks 和 L. R. Rabiner (“MPR”) 发表的一种经典方法，基于一种由俄罗斯数学家 E. Ya. Remez 改进的算法，在 DSP 领域中，该方法通常称为“雷米兹交换算法”。

专业视频中使用的高质量低通滤波器的系数如图 20-26 所示。

$$\begin{aligned} g[i] = & 0.098460s_{i-12} + \\ & 0.098482s_{i-11} - 0.013681s_{i-10} + \\ & 0.020420s_{i-9} - 0.029197s_{i-8} + \\ & 0.039309s_{i-7} - 0.050479s_{i-6} + \\ & 0.061500s_{i-5} - 0.071781s_{i-4} + \\ & 0.080612s_{i-3} - 0.087404s_{i-2} + \\ & 0.091742s_{i-1} + 0.906788s_i + \\ & 0.091742s_{i+1} - 0.087404s_{i+2} + \\ & 0.080612s_{i+3} - 0.071781s_{i+4} + \\ & 0.061500s_{i+5} - 0.050479s_{i+6} + \\ & 0.039309s_{i+7} - 0.029197s_{i+8} + \\ & 0.020420s_{i+9} - 0.013681s_{i+10} + \\ & 0.009482s_{i+11} + 0.098460s_{i+12} \end{aligned}$$

图 20-26 一个 25 抽头的 FIR 滤波器

20.17 信号重构

数字化过程涉及采样和量化。这些操作会在模/数转换器 (ADC) 中完成。信号是先量化后采样，还是先采样后量化，仅与 ADC 有关：操作顺序在子系统外是无关紧要的。现在的视频 ADC 是先量化后采样的。

前面已经解释过，在采样前需要进行滤波，这是为了避免引入混叠。在采样域中避免出现混叠是十分重要的。为了避免混叠，模拟预采样滤波器需要在模/数转换之前进行操作。如果避免了混叠，那么根据香农定理，采样信号就可以在无混叠的情况下重构。

要重构一个模拟信号，模拟重构滤波器必须跟在数/模转换之后，整个流程如图 20-27 所示。

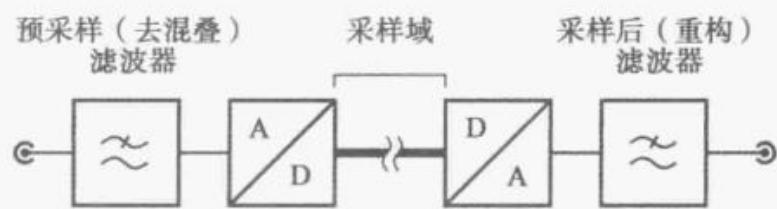


图 20-27 采样和重构

20.18 接近 $0.5f_s$ 的重构

考虑到图 20-28 所示的频率为 $0.44f_s$ 的正弦波的例子。该信号符合采样标准，可以

在数字域上完美地呈现。然而，从直观上看，由下图的采样点 3、4、5 和 6 很难预测出基本的正弦波。在重构时使用高斯滤波器，高频信号就会消失。为了进行准确的重构，如果波形在采样频率的一半附近占有很大的值，则必须使用高质量滤波器进行重构。

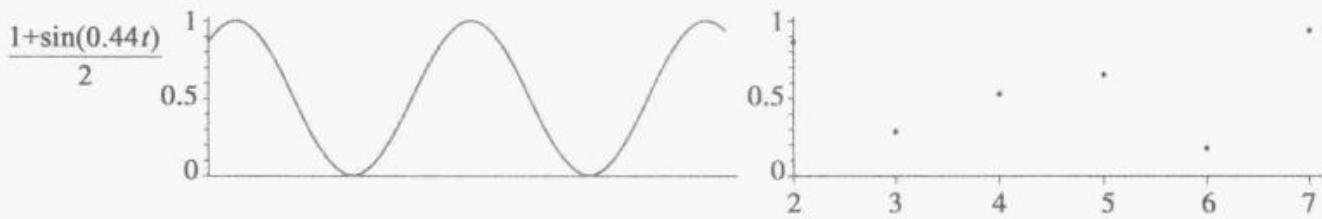


图 20-28 频率接近于 $0.5f_s$ 的重构

20.19 “ $(\sin x)/x$ ” 修正

前面已经解释过模拟重构滤波器跟在数/模转换之后的必要性。如果 DAC 产生冲激序列，其冲激的幅度由相应的代码值进行调整，一个经典的低通滤波器就可以适用：如果 DAC 输出类似于本书中的“点图”，即在采样时刻有值，非采样时刻无值，则是相当好的。回想一下，一个包含单位冲激的波形会在整个频谱上具有相同的频率响应。

这里给“ $(\sin x)/x$ ”加了引号，正确应为： $\sin(\pi x)/(\pi x)$ ，但是通常发音是 sine-ecks-over-ecks，将参数省略了。

对于模拟信号重构来说，不幸的是，典型的 DAC 无法为每一个采样点产生冲激波形。每一个冲激响应都有一个对应的 DAC，这是不现实的，因为信号的功率是与信号的积分成比例的，由于冲激信号的积分会实现足够的信号功率，所以冲激信号幅度会无限大。相反地，每一个转换完的采样值会持续整个采样过程：典型的 DAC 产生一个矩形波形。矩形波的频率响应用 $\text{sinc}(\cdot)$ 函数描述。

你可能认为 DAC 的矩形波形是“采样并保持”的操作，但其实该术语更常用于模/数转换的连接，或是 ADC 之前的电路。

图 20-29 所示的等同于数/模转换产生出一串脉冲序列后，用矩形滤波器的 $\text{sinc}(\cdot)$ 响应进行滤波。频率接近于 $0.5f_s$ 的信号是衰减的。在图 20-29 所示曲线中，顶部的曲线是频率为 $0.44f_s$ 的正弦波；底部的曲线是由传统的 DAC 产生的矩形波。尽管使用了高质量的重构滤波器，其响应延伸长度接近于采样频率的一半，但仍然很明显，使用矩形函数重建会降低信号高频分量的幅度。

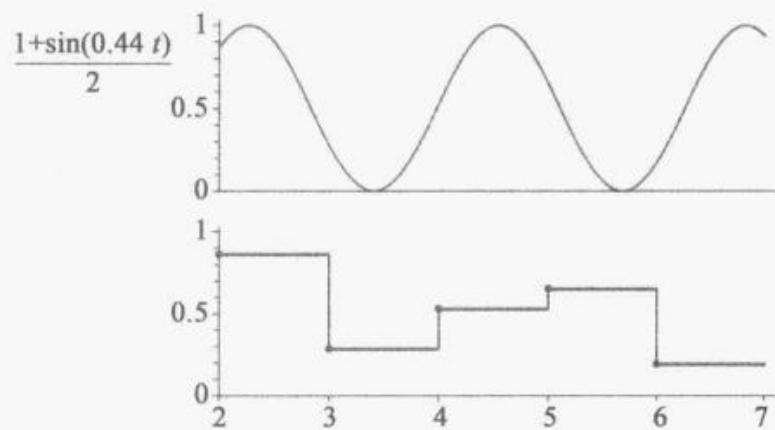


图 20-29 使用矩形波滤波的数/模转换

在采样间隔中每个采样值的 DAC 保持（零阶保持，Zero-Order Hold，ZOH）都对应着一个用 $\text{sinc}(\cdot)$ 函数的频率响应进行的滤波操作。图 20-30 所示的顶部曲线显示的就是由这种现象引起的衰减。

这种影响可以通过“ $(\sin x)/x$ ”来修正：修改重构滤波器的频率响应使得其可包含相应的 $\text{sinc}(\cdot)$ 函数倒数的峰值。在通带中，滤波器响应会比直流时的响应逐渐增长约 4dB，用于补偿损耗。当大于通带边沿频率时，滤波器响应必须迅速下降，以在采样率一半的频率附近产生较大的衰减，提供无混叠重建。图 20-30 显示了用“ $(\sin x)/x$ ”修正的理想化滤波器的响应。

本章详细介绍了一维滤波。第 22 章将会介绍二维或三维采样和滤波器。

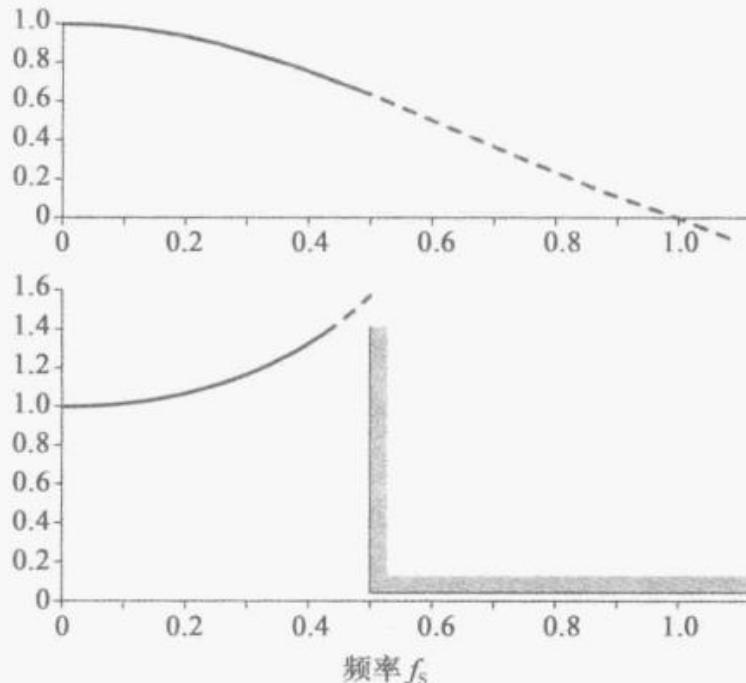


图 20-30 “ $(\sin x)/x$ ” 修正

延伸阅读

对于数字信号处理 (DSP) 的内容、理论, 以及数学方法更易于理解的介绍可参见 Lyons 的著作, 或者参见 Rorabaugh 的书, 后者涉及滤波器设计的源代码, 用于评估滤波器的系数。对于 DSP 综合和理论化的见解, 参见 Mitra 和 Kaiser 的著作。

Lyons, Richard G. (1997), *Understanding Digital Signal Processing* (Reading, Mass.: Addison Wesley).

McClellan, James H. and Parks, Thomas W. (2005), “A personal history of the Parks-McClellan algorithm,” *IEEE Signal Processing Magazine* 22 (2): 82-86.

Mitra, Sanjit K., and James F. Kaiser (1993), *Handbook for Digital Signal Processing* (New York: Wiley).

Rorabaugh, C. Britton (1999), *DSP Primer* (New York: McGraw-Hill).

第 21 章

重采样、插值和抽取

在视频和音频信号处理中，取一组采样值然后产生另一组在不同时刻以不同速率或不同相位产生的近似原采样值是很必要的。这就是重采样（PC 的术语中，用于调整图像大小的重采样称为缩放）。重采样是视频处理的核心步骤，其应用如下：

- 色度亚采样（例如 4：4：4 to 4：2：2）。
- 降频转换（例如 HD 到 SD）和升频转换（例如 SD 到 HD）。
- 宽高比转换（例如 4：3 到 16：9）。
- 视频标准中不同采样率之间的转换（例如 $4f_{SC}$ 到 4：2：2, 13.5MHz）。
- 在数字视频中图像调整的影响（DVE）。

一维重采样可以直接应用于数字音频中，例如，将采样率从 48kHz 变换到 44.1kHz。在视频中，一维重采样可以在水平方向或者垂直方向上使用。重采样可以扩展为二维采样序列，有两种方法可行。一种是先用水平滤波器，再用垂直滤波器，可以通过级联（串联）来实现——这是可分离的方法。或者，也可以通过直接进行二维空域内插实现。

上采样会产生多于输入样本的结果样本。在音频中，新的样本可以在高于输入样本的速率进行估计，例如，将 44.1kHz 的音频采样转换成 48kHz 的用于视频中的专业速率。在视频中，上采样要求从 1280×720 像素的 HD 转换成 1920×1080 像素的 HD：输入的每行 1280 个样本必须在输出中转换成 1920 个样本，上采样比率是 2：3[⊖]。

我在输入采样中写入重采样比率：输出采样。在我的习惯中，比例低于 1 的就是上采样。

实现上采样的一种方法为按 $1:n$ 的整数比在每对输入样本之间插入 $n-1$ 个零值。这会导致原信号的频谱以多倍原信号采样率的速率重复。重复的光谱称为“影像”（这是一个由无线电而来的历史名词，与图像并没有任何关系）！这些“影像”会通过消隐低通滤波器进行滤除（至少会将其衰减）。在一些上采样结构中，比如拉格朗日插值器，后续章节将会讲到，过滤和上采样是交织在一起的。

下采样获得少于输入采样的采样值。在音频中，新生成的采样的采样率要比输入采样的采样率要低。在视频中，在 $4f_{SC}$ NTSC 数字视频转换成 BT. 601（“4：2：2”）数字视频时要求使用下采样：每个输入行 910 个采样需要在输出中转换成每行 858 个采样，下采样率为 35：33；即对每 35 个采样输入，会有 33 个采样输出。

在原始采样序列中，从直流到频率为 $0.5f_s$ 的信号内容都可以呈现出来。经过下采样后，新的采样率会比要求的信号带宽要小。经过下采样后，有意义的信号内容在新的采样速率下受奈奎斯特准则限制——例如经过 4：1 的下采样后，信号内容会限制为原信号速率的 $1/8$ 。为了避免引入混叠，低通滤波器需要在下采样之前使用，或者与下采样结合使用。转角频率取决于下采样的比例；例如，采样比为 4：1 时需要转角角频率小于 $0.125f_s$ 。采样比为 $n:1$ 整数比的下采样可以认为是为新的采样速率而进行前置滤波（或

[⊖] 原书此处为 1080，有误。——译者注

平滑滤波)，是在原采样信号对之间丢掉 $n-1$ 个采样值。

重采样产生的新采样值被认为与其相邻采样有连续函数关系。如果给定函数不是连续的，可能会出现问题。例如，伪彩色图像不连续，无法在不失真的情况下进行有意义的重采样。

考虑对伪彩色数据进行重采样，如果将数据看作是连续的，那采样的结果可能会包含源信号中没有的颜色。如果使用最近邻域重采样法来避免产生新的采样值，则几何结构会被损坏。

图 21-1 显示上采样的频域。图 21-2 描述了原始信号的频谱。图 21-3 显示了下采样的频域。这些例子的显示比例是 1:2 和 2:1，不过本例的概念适用于任何比例的重采样。

图 21-1 通过在原采样对中插入零采样值进行的 2 倍上采样。这会导致源信号的折叠谱的内容以新的速率出现在带宽内。这些“影像”通过重采样滤波器进行移除。

图 21-2 源信号在临近半采样率时会发生折叠。这对信号能否被正确重建是无关紧要的。当信号进行上采样或下采样时，折叠部分必须合理处理，否则会出现混叠。

图 21-3 所示的是 2:1 的下采样，需要一个重采样滤波器以新的采样速率来满足奈奎斯特准则。深色实线表示的是滤波后信号的频谱；虚线表示折叠部分。未经滤波的重采样会保留原始基带信号的频谱。但是在新的采样速率附近折叠会导致混叠产生，如图中交叉线阴影所示。

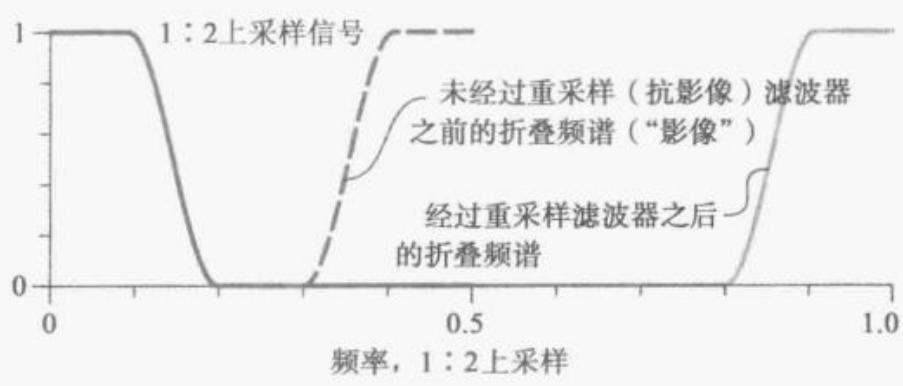


图 21-1 两倍上采样

上采样

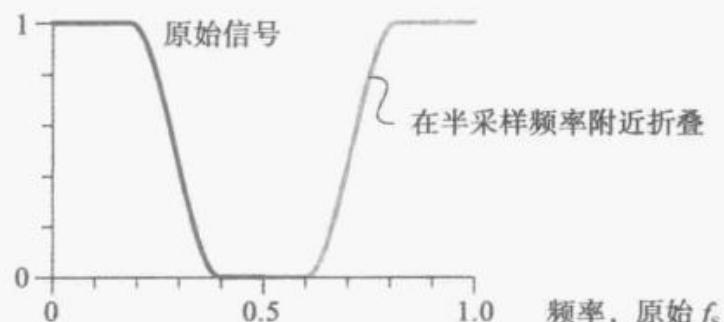


图 21-2 源信号

下采样

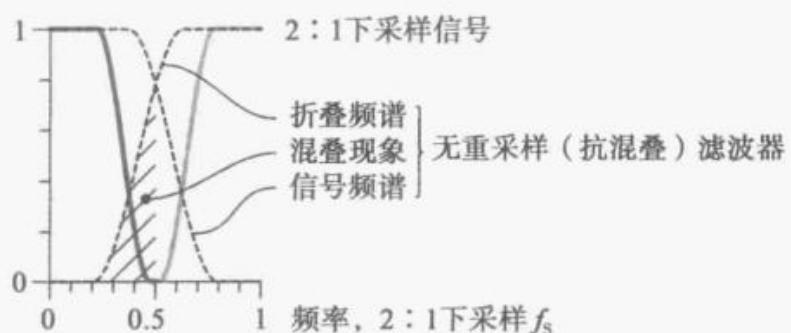


图 21-3 2:1 的下采样

21.1 2:1 下采样

彩色视频来源于 $R'G'B'$ 分量。如果信号是用于演播室的，则转换成 $Y'C_B C_R$ 是必要的。转换包括矩阵变换（转成 4:4:4 格式的 $Y'C_B C_R$ ），然后色度亚采样成 4:4:2 格式。色度亚采样需要一个 2:1 的下采样器。如果打算通过简单丢弃交替样值来实现下采样，则频率在 $0.25f_s$ 和 $0.5f_s$ 之间的信号都会引起混叠。频率高于 $0.25f_s$ 的信号内容必须丢弃。这里需要的滤波器通常是转角频率为原始采样率 $1/4$ 的 FIR 低通滤波器。滤波之后，交替结果采样值可能会被丢弃。对后续要丢弃的值进行计算没有必要，不过，高效的色度亚采样正是利用这一事实的优势，将 C_R 和 C_B 分量插入单个滤波器中的。

图 20-12 给出一个很简单的低通滤波器，对相邻采样值求简单平均值。滤波器的转角

频率为 $0.25f_s$ 。不过，它的通带和阻带之间具有较宽的过渡带，且阻带的截断效果较差（高于 $0.25f_s$ ）。它是较差的重采样滤波器。滤波器的抽头要求多于两个才能为演播室的视频下采样提供足够好的性能。

在 4:2:2 视频中，色度和亮度是同址的，每个色度采样必须位于亮度采样的点上。具有偶数个抽头的对称滤波器不具有这个属性。色度同址的下采样滤波器必须具有奇数的抽头。

21.2 过采样

前面已经解释过在模/数 (A/D) 转换之前进行前置滤波和在数/模 (D/A) 转换之后进行后置滤波的重要性。以前，这些滤波器都是在模拟域中用电感和电容实现的。在离散的形式中，这些元件笨重且昂贵。将具有合适的数值和精度的电感和电容元件集成到集成电路中是相当困难的，不过，模/数转换器和数/模转换器的工作速度越来越高，数字运算变得非常廉价。这些情况导致了过采样作为一个经济划算的选择代替复杂的模拟预采样（“抗混叠”）和后置采样（重建）滤波器。

传统的模拟预采样滤波器的特性是十分关键的：在低于 0.4 倍采样速率时衰减必须很小，高于 0.4 倍采样率时衰减要很高。在演播室视频的预采样滤波器中，在低于 0.5MHz 的频段中衰减低于 1dB 或最高达到 5.5dB，在高于 6.75MHz 时衰减最好大于 40dB 或 50dB。这就要求转换比率为 $\Delta\omega/\omega_S$ 。图 21-4 画出了传统的模拟预采样滤波器的模型（由于这种滤波器必须满足严格的约束条件，这使其成本变得十分昂贵）。

过采样模/数转换器工作在多倍极限采样速率下——如 27MHz，为 BT.601 视频采样率的 2 倍。信号在经过廉价的模拟滤波器之后经过该转换器，模拟滤波器对 13.5MHz 及高于该频率的信号分量有很强的衰减作用。不过，它在 5.5MHz 和 13.5MHz 之间频段的特性并不好。在该频段的滤波器需求留给了数字 2:1 下采样器，模拟滤波器的转换比 $\Delta\omega/\omega_S$ 与直接转换的比值已大大放宽。以现在的技术，数字下采样器的花费低于优秀和中等模拟滤波之间的花费。复杂度从模拟域转移到数字域；总的系统损耗减少。图 21-5 画出适用于 2 倍过采样模/数转换器之前使用的预采样滤波器的模型。因为滤波困难，该滤波器被转移到数字域上，所以其要求比运用于直接采样的滤波器要少很多，获得的频率响应如图 21-4 所示。

图 21-5 显示的是转角频率为 $0.25f_s$ 的 55 抽头滤波器响应。这是一个半带滤波器，用于 2 倍过采样模/数转换器之后。

某些 FIR 滤波器的转角频率正好是 $0.25f_s$ ，一半的系数为零。这导致复杂度大大降低。

这一处理过程与 2 倍过采样数/模转换器的类似，数/模转换器工作在 27MHz 上，其呈现的数据流是经过 1:2 上采样的。对于每个输入采样，经过采样滤波器计算出 2 个输出采样。其中一个是输入采样的

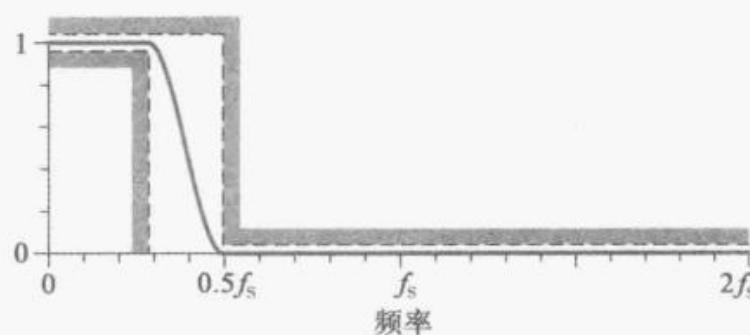


图 21-4 用于直接采样的模拟滤波器

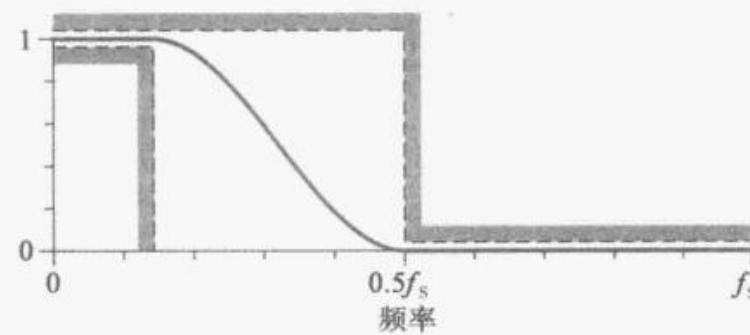


图 21-5 一个用于 2 倍过采样的模拟滤波器

有效位置上进行计算的，另一个是在输入信号之间的一半有效位置上进行计算的。该滤波器在 6.75MHz 和 13.5MHz 之间功率有衰减。模拟后置采样滤波器只需要丢掉高于 13.5MHz 的分量。与 2 倍过采样模/数转换器采样过程一样，其在 6.75MHz 到 13.5MHz 之间的性能并不好。

21.3 插值

在数学上，插值是计算函数或者假定函数值（称为 \tilde{g} ）的过程，对于一个任意的参数 x ，给定一些函数自变量和数值对 (x_i, s_i) 。插值的方法有很多，构造插值函数的方法也有很多。

在常见的插值穿过图像行的例子中，参数 x 是水平位置，沿着时间轴插值，如在数字音频采样率转换时，可以使用符号 t 表示时间。

给定了两组采样值对 (x_0, s_0) 和 (x_1, s_1) ，其线性插值函数可以写成：

$$\tilde{g}(x) = s_0 + \frac{x - x_0}{x_1 - x_0} (s_1 - s_0) \quad (21-1)$$

这里使用 \tilde{g} 来表示插值函数， f 表示频率。用 g 加上波浪线 (\tilde{g}) 来强调它是个近似值。

在计算机图形学中，线性插值运算称为 LIRP (读作 lerp)。

线性插值函数也可以用相邻采样 s_0 和 s_1 的加权形式表示为：

$$\tilde{g}(x) = c_0(x) \cdot s_0 + c_1(x) \cdot s_1 \quad (21-2)$$

其权重由 x (或者 t) 的坐标确定：

$$c_0(x) = \frac{x_1 - x}{x_1 - x_0}; \quad c_1(x) = \frac{x - x_0}{x_1 - x_0} \quad (21-3)$$

21.4 拉格朗日插值

J. L. 拉格朗日 (1736—1813 年) 发明了一种使用多项式的插值方法。三次插值函数的多项式形式为：

$$\tilde{g}(x) = ax^3 + bx^2 + cx + d \quad (21-4)$$

插值法包括基于给定的参数值对 (x_j, s_j) ，选择合适的参数 a 、 b 、 c 和 d 。拉格朗日描述了一种简单、便捷计算参数的方法。线性插值只是一阶拉格朗日插值法的特例（直接使用最近邻域值被认为是零阶插值）。也有二阶（二次方程式）形式，但是很少用于信号处理。

在数学上内插就是指前面所描述的过程。然而，相同的词可用来表示由插值函数生成的正好在原采样位置 x_i 上的原采样值 s_i 的特性。拉格朗日函数表现了该特性。读者可能会猜想这一特性是任意插值函数都需要的。不过，信号处理中并没有这一要求，事实上，用于视频和音频的插值函数极少通过原采样值。使用数学术语的结果就是，在视频中，我们有着看似自相矛盾的情况，插值函数通常并不“插值”。

原理上，三次插值可以使用任意参数 x_i ，即使该值在四个输入采样值 x 坐标的范围之外（在区间 (x_{-1}, x_2) 上的估计称为外插法）。在数字视频和音频中，将 x 的值限制在 x_1 和 x_2 之间，以便估计在中心两个采样点间隔之间的信号。要估计该区间之外的值，适当替代输入采样值 (s_{-1}, s_0, s_1, s_2) ，例如，要在 s_1 和 s_2 之间进行估计，则将输出采样值向左移动一个位置。

使用均匀采样（如在传统的数字视频中），在两个中心采样之间插值时，参数 x 可以

改写成相位偏移，或者分数相位 (φ , phi)，其中两个中心采样之间需要一个新的采样值（见式 (21-5)）。在抽象术语中，是介于 0 和 1 之间的；在硬件上，使用二进制数或有理分数来实现。在视频中，一维插值器通常为 FIR 滤波器，其系数为相位偏移的函数。权重系数 (c_i) 是相位偏移量的函数；可以看成是基函数，即

$$\varphi = \frac{x - x_0}{x_1 - x_0}; x_0 \leq \varphi \leq x_1 \quad (21-5)$$

在信号处理中，三次（三阶）插值经常使用，该情况如图 21-6 所示。在线性插值中，需要用到左右的相邻值。在三次插值中，使用预插值样本左侧和右侧各两个原样本值，在中心位置进行插值。

式 (21-2) 可以改写成：

$$g(\varphi) = c_{-1}(\varphi) \cdot s_{-1} + c_0(\varphi) \cdot s_0 + c_1(\varphi) \cdot s_1 + c_2(\varphi) \cdot s_2 \quad (21-6)$$

该方程在待插值的采样点位置附近取四个采样值 (s_{-1}, s_0, s_1, s_2)，相位偏移为 0 到 1。系数 (c_i) 是当前参数的函数；插值器得到四个采样值的加权和值，其中权重是参数 φ 的函数，返回一个估计值（如果输入采样值是一个没有超过三次的多项式的值，则该值可以由三次拉格朗日插值器生成，其中包含舍入误差：拉格朗日插值的“插值”）！

如果二维图像阵列在任意的 x 和 y 坐标点上重采样，一种方法是用一维滤波器

沿着一条坐标轴进行滤波，再沿另一条坐标上进行滤波。这种方法将插值作为一个分离的过程，类似这种分离滤波的方法将会在后续章节中进行介绍。令人感到惊讶的是，这种两步走的方法可以用来旋转图片（史密斯所著）。或者，每一步用一个 2×2 的阵列（四个采样值）进行二维线性插值——即双线性插值。一个更加复杂的方法是，使用 4×4 阵列作为三次插值的模板进行二维插值——即双三次插值（它在数学上可以媲美一维 15 度插值）。

曲线可以用参数 u 在二维空间中绘制出来，两个函数 $x(u)$ 和 $y(u)$ 的参数为每个 u 值生成一个二维坐标对。该方法可以通过增加第三个函数 $z(u)$ 从而扩展到三维空间中。贝塞尔（Pierre Bézier）发明了一种已广泛使用的新方法，运用三次多项式来描述曲线和曲面。这样的曲线现在称为贝塞尔曲线。该方法在计算机图形领域非常重要，不过，贝塞尔曲线和其类似曲线在信号处理中运用很少。

21.5 用于滤波的拉格朗日插值

除了用 4 抽头代替 5 抽头，式 (21-6) 与式 (20-2) 的 5 抽头高斯滤波器几乎是一样的。拉格朗日插值器可以看成是特殊的 FIR 滤波器，故可以将其作为滤波过程分析。在之前的第 20 章中，所举的例子都是对称的。在输入采样一半间距的位置产生采样的插值也是对称的，例如，2 倍过采样数/模转换。但是，大多数插值器是非对称的。

只有对称的 FIR 滤波器才有真正的线性相位，其他的滤波器只是表现性能接近线性，接近到在视频和音频中可以近似认为是线性的。

多项式插值通常不适合视频信号处理有四个原因：多项式插值有着不相等的阻带纹

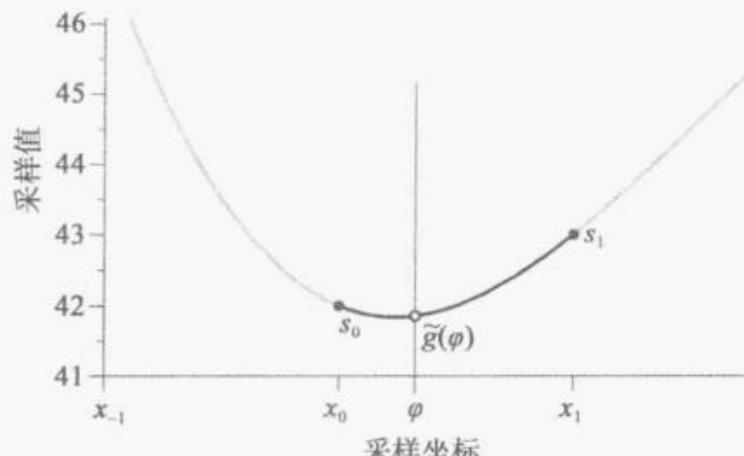


图 21-6 三次插值

波；空白段位于阻带的固定位置；插值函数显示了中心间隔之外的过激行为；由插值器产生的信号会存在一些噪声。以下将针对每一个问题加以说明。

- 任何拉格朗日插值的频率响应，其阻带纹波都不相等，有时候极度不稳。这通常是在信号处理中不希望出现的，在视频中也不期望出现。
- 拉格朗日插值器在原采样点间进行插值；这导致了幅频响应出现周期性的空白段（零值），其频率由插值器的阶数确定。滤波器设计者要控制阻带衰减，就需要自由地设置这些空白段。这种自由不适用于拉格朗日插值器的设计。
- 从概念上讲，插值过程尝试使用相关的简单函数对生成采样值的未知函数进行建模。使用的函数形式需要反映信号产生的过程。一个三次多项式可以实现两个中心点之间的合理内插。不过，任意多项式的值在原采样值限制的区域外可以迅速达到正无穷或者负无穷。该性质与信号的动作不一致，信号的值永远限制在有限的范围内（在视频中是 0~1，在音频中是 0~0.5）。
- 在信号处理中，在采样值中总有一些由信号的噪声、量化噪声和数字域中计算时的舍入误差造成的不确定性。当源数据是不完美时，要求插值功能完美似乎是不合理的。

在信号处理中提出的这四个问题可以通过非多项式插值函数和非经典数学插值函数来解决。相反，本书经常使用基于 $\text{sinc}(\cdot)$ 加权函数的插值函数，见 20.5 节。在信号处理时，经常设计插值器使其不对原采样进行插值。

理想的 $\text{sinc}(\cdot)$ 加权函数在频谱上没有明显的空白段，当 $\text{sinc}(\cdot)$ 函数被截断并优化以获得一个物理可实现的滤波器时，阻带中的零值数目有限。不同于拉格朗日插值器，这些零值不是等间隔的。为零点选择合适的频率，以调整滤波器的响应是滤波器设计者的权利。

可以考虑使整个理想 $\text{sinc}(\cdot)$ 滤波器的阻带包含无限个零值。在数学上， $\text{sinc}(\cdot)$ 函数用多项式的阶数表示拉格朗日插值接近于无穷。参见史密斯的《Digital Audio Resampling》附录 A。

21.6 多相插值器

一些视频信号处理需要以简单比率进行上采样，例如，在高清格式转换由 1080 像素转成 1920 像素需要 2 : 3 上采样。一个输出采样在三分之一相位处进行计算。不管是在输入采样点的位置，或者在输入采样之间以 1/3 或 1/2 处，上采样器可以仅通过三组参数以 FIR 滤波器来实现，系数可以通过 φ 查找表来获得。

720p60 帧和 1080i30 帧具有相同的采样率 (74.25MHz)。在本例的逻辑设计中，为单时钟域。

许多插值器中涉及的比率会比上例中的 2 : 3 的比率更复杂，例如，由 $4f_{\text{SC}}$ NTSC 转换到 BT. 601 (4 : 2 : 2) 时，910 个输入采样必须转换成 858 个采样，它的下采样比率为 35 : 33。连续的输出采样是通过对输入采样进行 12/33 上采样计算得来的。每一个第 33 个采样在输入采样 (0) 处计算。其他采样值分别在输入采样坐标位置的 12/33, 24/33, …, 1632/33, 181/33, 193/33, …, 3431/33 处。在每一个输出采样点，寻址电路需要一个递增的样本计数器，递增值为 1 加上分数 2 模 33 (得到分数 2/33)。从分数计算器到采样计数器转换过程中的溢出值，在本例中占了采样序列中 17 个输入采样数。所需的插入相位为输入采样点间的 0, 1/33, 2/33, 3/33, …, 32/33。

一种直接在硬件上设计插值器的方法是，以输入采样率驱动一个 FIR 滤波器。在每一

个输入时钟，输入采样值在寄存器间移位。寻址电路使用模 33 计数器来追踪相位——即相位累加器。每一个时钟，将有 33 组不同系数中的一组系数应用到滤波器上。每组系数设计时都引入合适的相位偏移。在该例中，每 35 个输入时钟仅要求 33 种采样结果：每 35 个时钟中有两个时钟不产生结果。

本例的逻辑设计包含了两个时钟域。

这种结构称为多相滤波器。该例包含 33 种相位，不过，需要的抽头数量与相位数目无关。一个两倍过采样预滤波器也只有两个相位（见 21.2 节）。半带滤波器的响应如图 20-25 所示，也适合于这种应用，该滤波器有 55 个抽头。

21.7 多相抽头和相位

滤波器中所需抽头的数目取决于设计者对频率响应控制的程度，也取决于每个相位与其他相位匹配的紧密程度。在很多消费级视频中，三次（4 抽头）插值足够使用了。在演播室视频中，滤波器必须有 8 抽头或者更多，这取决于要达到的性能目标。

在多相 FIR 插值器的直接实现中，相位的数目取决于相关采样率的算法。相位的数目决定所需使用的系数的数目。系数集通常预先计算出来，并存储在非易失性存储器中。

21.6 节讲到了一种拥有 33 个相位的多相位重采样器。在一些应用中，相位的数目大到无法直接实现，比如从 $4f_{SC}$ PAL 转换成 BT. 601 (4 : 2 : 2) 需要 $709379 : 540000$ 的转换率，从每行 922 个有效采样转换成 702 个。在其他应用，比如数字视频效果中，相位的数目是变量，而且不能提前预知。这类应用可以通过使用具有相位数为 2 的幂次的插值器来寻址，如具有 256 个相位。相位偏差按合适的精度来计算，可以近似一个二进制分数（在该例上是 8 位）来为插值器得到相位偏移。

如果插值器要实现 8 分位数，则算得的输出采样可能会产生至多 $\pm 1/512$ 的采样间隔误差，这对于数字视频是一个完全可以接受的误差。不过，如相位累加器实现的仅是 8 分位，位置误差会随着图像行增量计算过程而累加。在该例中，每行有 922 个有效采样，误差可能会出现在每行末端 3 到 4 个样本间隔处！这是不可忍受的。解决方法是选择一个足够的分位数使得累计误差不超过容许范围。在该例中，13 位足够大了，但只要把其中 8 位送入插值器中。

21.8 多相插值器的实现

多相插值是 FIR 滤波器的特例，但是在实现上主要有三点不同。第一，在通常的 FIR 滤波器中，输入和输出速率是相同的；在多相插值器中，输入和输出速率通常是不同的。第二，FIR 滤波器通常有固定的系数；在多相 FIR 插值器中，系数随采样基的变化而变化。第三，典型的 FIR 滤波器是对称的，而多相插值器不是对称的。

一般来说，对于相位数目较少时，如 8 个或更少，插值器的开销是受乘法运算的数量控制的，乘法运算的数量与抽头数成正比。超过 8 抽头后，系数存储的开销开始显著体现出来。寻址电路的开销值仅取决于相位的数目。

在 35 : 33 下采样例子中，已经讨论了由输入采样速率驱动的硬件结构。假设硬件设计需要由输出时钟驱动插值器。对于每 33 个输出时钟中的 31 个时钟，都消耗一个输入采样；不过，剩余的 2 个时钟，则每个时钟消耗两个输入采样。这对存储系统设计加上了一个限制：要么存储器中必须实现双通道，要么每行增加额外的 44 个采样在空间隔中通过，然后存储在一个小缓存器中。通过输入时钟控制插值器要更容易一些。

考虑一个从 BT. 601 转换到 $4f_{SC}$ NTSC 格式的 33 : 35 上采样器。如果从输出端驱动，

则插值器每个时钟都会产生一个输出采样，每个时钟最多消耗一个输入采样（35 个时钟中有两个时钟，没有消耗输入采样）。如果从输入端驱动的话，对于 33 个输入时钟中的两个时钟，插值器必须生成两个输出采样。这些在 FIR 滤波器和输出端存储系统设计中很可能出现问题。

从中可以得到经验：若是由高速端驱动，则多相插值器的结构就可以得到简化。

21.9 抽取

在拉格朗日插值中，不会考虑插值计算得到输出采样是比输入采样多还是少。不过，在信号处理中，下采样和上采样会有很大的不同，下采样时必须使用低通滤波器防止混叠；在上采样时，必须用低通滤波器抑制出现“镜像”。在信号处理中，插值通常是指上采样。也即以单位比率或更高的比率来重采样（插值也可以描述采样率不变时的相位变化，可以将其认为是采样率为 1:1 下的特殊例子）。

10:9 的下采样类似于罗马军队对待背叛和兵变士兵的政策：将十分之一犯事的士兵处死，这里的抽取通常被描述成下采样。

按字面上意思理解，抽取比例是 10:9，而不是 10:1。

21.10 用于抽取的低通滤波器

在本章的前面，曾讲述了 2:1 抽取的色度亚采样。在抽取器中，采样值以新采样率的一半或者更高的速率进行低通滤波，以衰减分量，然后将采样值丢弃。显然，采样值将会直接丢弃而无需经过计算。通常情况，采样丢弃和滤波在同一电路中进行。

在色度亚采样的半带抽取的例子中，解释了低通滤波截止到 $0.25f_s$ 的必要性，在 $4f_{sc}$ NTSC 到 BT. 601 格式转换的例子中介绍了多相插值器，输入和输出速率很接近，所以不需要格外关注得到的采样率的带宽限制。如果下采样比率要远大于 1，比如 5:4，或者更高，则脉冲响应必定包含低通滤波（预滤波或者抗混淆）函数和相位偏移。为避免混淆，低通转角频率必须随下采样比率匹配缩放，这可能需要几组具有不同转角频率的滤波器系数组。

插值和抽取的详细内容，请参见 Crochiere, Ronald E. 和 Lawrence R. Rabiner (1983 年)《Multirate Digital Signal Processing》(New York: Prentice-Hall)。

第 22 章

图象数字化与重构

以足够高的速率对静止图像的序列进行捕捉和显示——通常每秒 24~60 帧图像——这样可以产生运动的错觉。在时域进行采样时，结合二维（空域）采样，使得数字视频在三个轴上进行采样：水平轴、垂直轴和时间轴，如图 22-1 所示。

一维采样理论见第 20 章。

一维采样理论分别运用在三个轴的每一轴上。图 22-1 只画出三个时间采样点，因为时间采样数受提供的画面数目限制。画面存储会比行存储代价更高，图中画出了 5 个垂直采样：每个垂直采样都与行存储有关。

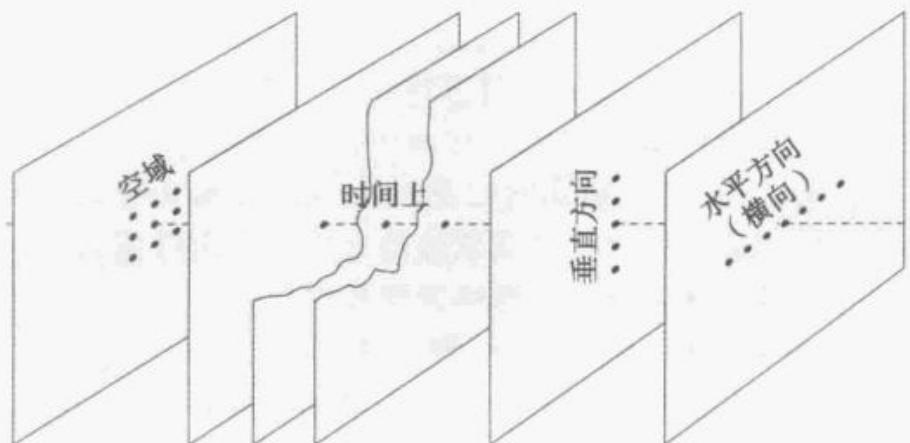


图 22-1 时空域

图 22-1 最左侧所示的是单个图像的二维空间域示意图。一些图像处理操作，比如某些滤波器，可以在水平轴和垂直轴分别执行，且对整个空域有效，这一操作过程称为分离。其他处理操作无法分离成垂直面和水平面操作，并且必须通过二维采样阵列直接执行，这时适用二维采样理论。

第 20 章描述了如何分析一维时间函数的信号，例如音频信号。采样理论也可应用于一维空间函数的信号，例如视频信号的单次扫（图像行）描行。这是水平域或横向域，如图 22-2 所示。如果对一个图像进行逐行扫描，那么每行的波形可当作独立的信号。在第 20 章中所讨论的一维滤波和采样技术，可以直接用在这里。

考虑到一组垂直的点在几个连续的图像行中处于相同的位置，如图 22-3 所示。这些点是扫描过程本身采样得来的。采样定理可用来理解这些采样的属性。

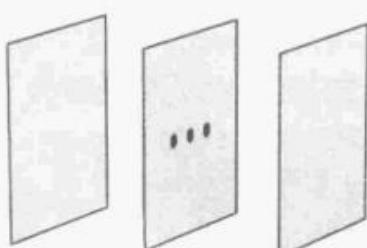


图 22-2 水平域

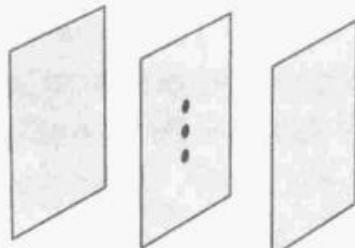


图 22-3 垂直域

当对连续图像在时间上采样来表现运动时，引入了第三个维度。图 22-4 描绘了三个连续帧中位于同一列和同一行的采样。

复杂的滤波器可以同时对两个轴进行操作，图 22-5 给出了空域采样示例。完整的采样集属性可以全部一起考虑，而无需按水平和垂直两个方向分别讨论。

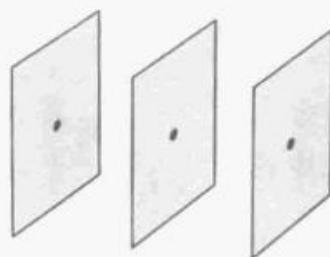


图 22-4 时间域

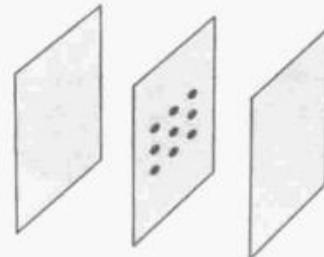


图 22-5 空间域

22.1 空间频域

本书在图像结构中已经讲过一维时域波形如何转换成一维频域信号。这一概念可以扩展到二维空间中，二维空间信号可以转换为二维空间频率。图像的内容可以用水平轴和垂直空间频率分量来表示。空间频率使用周期数/图像宽度 (C/PW) 作为 x 轴坐标单位，周期数/图像高度 (C/PH) 作为 y 轴坐标单位来绘制。通过研究空间频率响应来深入了解成像系统的操作。

在图 22-6 左上角所示的图像中，每个图像行都有相同的内容：4 个周期的正弦波。在图片下面，描绘了每一行的时间域波形。因为每一行都是相同的，所以在竖直方向不存在功率。从频率空间考虑，该图像功率包含在单个水平空间频率上的功率， $4\text{C}/\text{PW}$ ；在纵轴上没有任何功率。该图像的所有功率都位于空间频率 $(4, 0)$ 处。

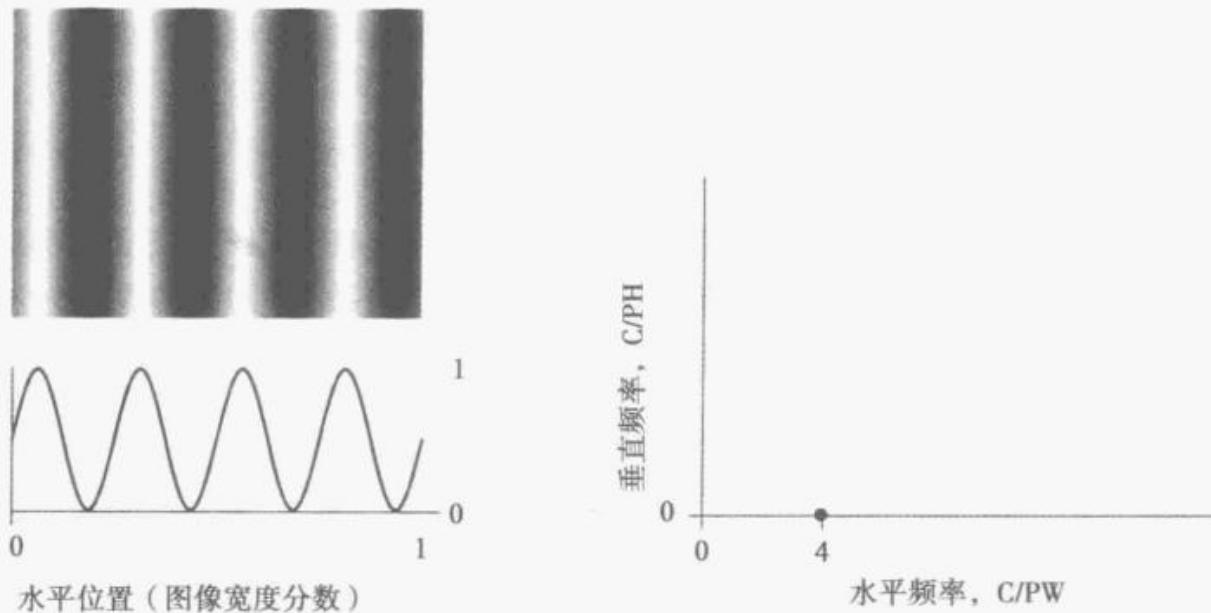


图 22-6 水平空间频率域

图 22-7 显示的是在垂直方向包含正弦波信号的图像。图像的高度包含三个周期。右侧的空间频率图，显示该图像的所有功率都包含在空域频率的 $(3, 0)$ 坐标处。若图像中每一行都为固定值，则所有的功率都位于空域频率的 y 轴上。

如果图像包含具有相同内容的行，则所有的功率都会集中在空间频域的水平轴上。如果连续扫描行的内容略有不同，功率会扩散到非零的垂直频率上。对角线图像在空间频域中会占据单个点，偏离 x 轴和 y 轴。

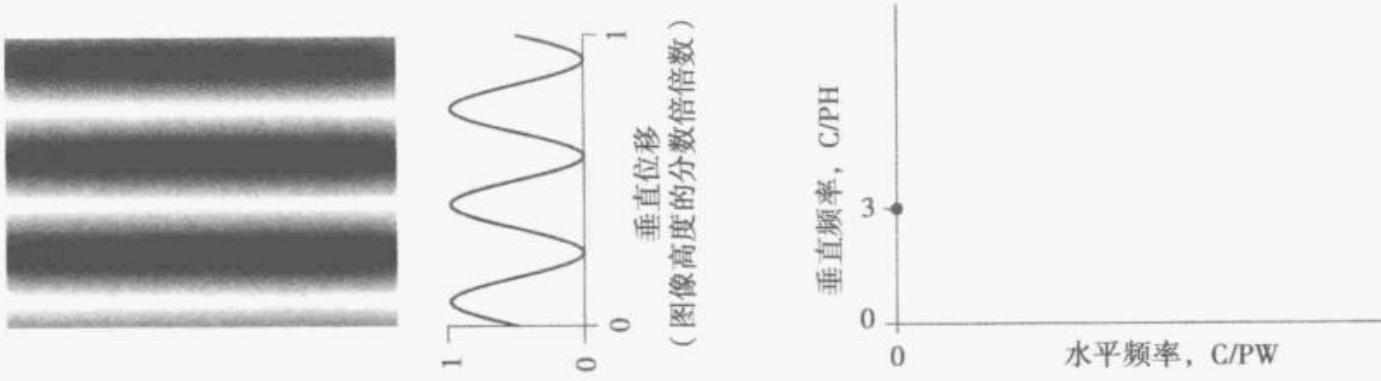


图 22-7 垂直空间频率域

对应于垂直采样率一半的空间频率取决于图像的行数：480i 系统具有约 480 个图像行；480 个采样就是图像的高度，垂直采样的奈奎斯特频率是 240 C/PH。垂直采样若超过该值，则在重现时就会发生混叠。

当空间频率通过二维傅里叶变换确定时，结果如图 22-8 所示。其中低垂直频率，也就是低 y 值显示在底部。当空间频率通过离散变换计算数值时，例如二维离散傅里叶变换 (DFT)、快速傅里叶变换 (FFT)、离散余弦变换 (DCT)，结果通过矩阵来表示，这里垂直低频部分位于顶部。

在大多数图像中，连续行和连续列的采样 (R' 、 G' 、 B' 、或者是色度) 值是非常相似的；低频是主要部分，图像功率主要集中在空域频率坐标 $(0, 0)$ 处。图 22-8 描绘了 480i 系统的亮度的空域频谱。如果一个未调制的色彩副载波是图像数据信号，则它将会出现在指定的位置。在复合 NTSC 中，色度会调制成副载波，产生的调制色度可以认为是在空间频域中占据了一定的位置，如第六章：传统视频系统的 NTSC 和 PAL 调制，在 NTSC 空间频谱中所讲。在 NTSC 编码中，调制的色度和亮度结合，这会导致频谱叠加。如果色度谱和亮度谱重叠，会出现交叉颜色和交叉亮度干扰效应。

在光学上，幅频响应和带宽等术语都用不到。一种光学元件、子系统或者系统的特点为其调制传递函数 (MTF)，一维水平或垂直的空间频率响应图（调制的深度为该图中引用的单个点）。技术上来说，MTF 是点分布函数 (Point Spread Function, PSF) 或行分布函数 (Point Spread Function, LSF) 的傅里叶变换。通过定义，MTF 与光的强度关联起来。因为负的光功率物理上无法实现，MTF 通过将高频正弦波叠加（调制）到一个恒定水平来测量。然后将输出调制的比率运用到输入调制中。

光学传递函数 (OTF) 中包括相位。OTF 的幅度是 MTF 的幅度，而 MTF 与相位无关。

22.2 梳状滤波

有限脉冲响应 (FIR) 滤波器中，已经讲到单一时间维上 FIR 滤波器的操作过程。如果采样是图像上的扫描线，则可以认为频率响应能用水平空间频率（单位是 C/PW），而不用时间频率（周期/s，或 Hz）来表示。

考虑一个来自数字图像序列的采样和紧挨其下方的采样，如图 22-9 所示。如果图像每行有 640 个有效采样，将这两个采样送入梳状滤波器，如图 20-19 所示，而在两个“1”

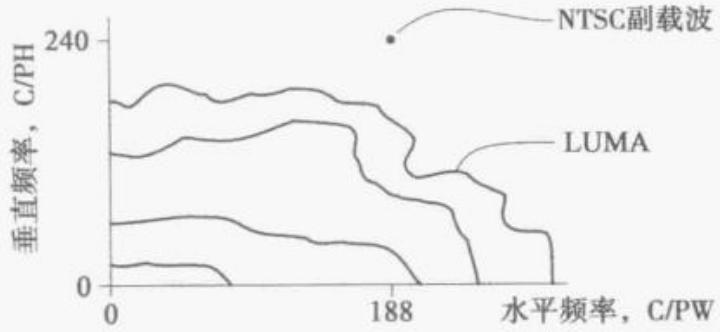


图 22-8 480i 亮度的空间频率频谱

之间有 639 个零采样。则梳状滤波器的操作相当于在垂直方向上有两个抽头权重为 [1, 1] 的滤波器。图 20-10 画出一个一维 [1, 1] 滤波器的频率响应，其中曲线显示了梳状滤波器的响应，表示了其在垂直方向上的响应。这里幅度响应关于直流增益进行了归一化，滤波器在四分之一垂直采样频率时约有 0.707 倍响应（也就是 3dB 衰减）。图 22-10 所示的为两行 (1H) 梳状滤波器，幅度衰减随 $\cos\omega$ 变化而变化。

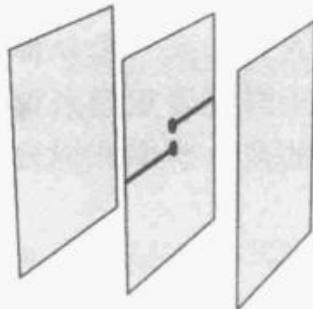


图 22-9 两个采样垂直放置

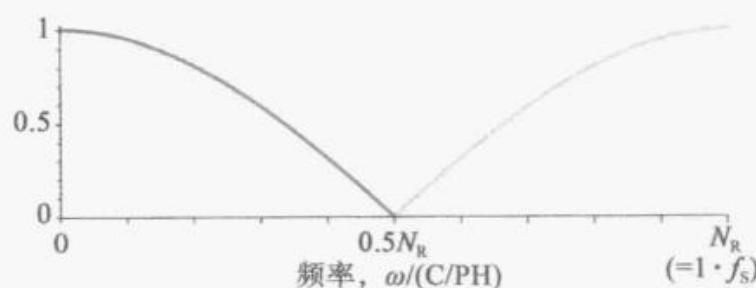


图 22-10 [1, 1] FIR 滤波器的响应

22.3 空域滤波

[1, 1] 水平低通滤波器与 [1, 1] 垂直低通滤波器串联（级联）等效于将如图 22-11 所示的左侧的矩阵作为权重系数计算加权和。一个 [1, 2, 1] 水平低通滤波器与一个 [1, 2, 1] 垂直低通滤波器串联相当于将图 22-11 所示的右侧的矩阵作为系数计算空间采样的加权和。这就是空域滤波器的例子。这些特定的空域滤波器是可分离的：它可以通过水平滤波器和垂直滤波器串联实现。许多空域滤波器是不可分离的：它们的计算必须是在二维空间域中直接进行；它们无法使用一维的水平滤波器和垂直滤波器级联实现。不可分离滤波器的例子如图 22-12 所示。

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

图 22-11 可分离空间滤波器的例子

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

图 22-12 不可分离空间滤波器的例子

22.4 图像预采样滤波器

在相机中，连续的信息都需要经过预采样（抗混叠）滤波。混叠可通过光学空间低通滤波器在图像信号转换为电信号之前影响光学路径来减到最小。棱镜的 MTF 限制是强加一定程度的滤波。辅助滤波器可以看作独立的光学器件来实现（通过使用双折射的光学理论）。此外，某种程度的滤波可以通过光电传感器本身的光学特性加强。

在重采样中，信号功率并不局限于正值，滤波器也可以使用负的权重。ILPF 和其他正弦基滤波器都具有负的权重值，但这些滤波器经常产生振铃现象或表现出较差的视觉性能。Schreiber 和 Troxel 发现精心设计的 $\sigma=0.375$ 的锐化高斯滤波器比 ILPF 具有更好的性能。经过特定数学准则优化后的滤波器并不一定会产生视觉效果最好的图像！

22.5 图像重构滤波器

本书曾介绍过“矩形框滤波器”重构，在技术上这称为“采样/保持”“零阶保持”或

“最近邻域”重建。

从理论上来说，理想图像重构可通过使用具有二维的 $\text{sinc}(\cdot)$ 分布的 PSF 函数来实现。这是一个二维的 ILPF，之前介绍过一维的 (ILPF)。然而， $\text{sinc}(\cdot)$ 函数含有负的偏移。光功率不能为负数，所以 $\text{sinc}(\cdot)$ 滤波器无法在图像捕捉设备中用于预采样，也无法在图像显示器中用于重构采样。传感器的每个元素灵敏度的矩形分布很容易实现。然而像第 20 章所述的一维矩形框，矩形分布在高频段有明显的响应。矩形框用在传感器上可以避免混叠。用在显示器上，扫描线或像素结构可能是可见的。如果一个外部光学器件例如镜头对空间高频段衰减，则矩形分布可能合适。对于图像采样或者重建来说，选择具有半功率带宽的高斯函数更为简单实用。高斯函数是实现合理的高分辨率和混叠最小化和像素结构（或扫描线）可视度最小化的折中考虑的结果。

升余弦分布类似于高斯分布，Schreiber 和 Troxel 建议采用 $\sigma=0.3$ 的锐化高斯分布。

22.6 空间（二维）过采样

在图像采样与图像显示的重建中，理想的理论性能可以通过使用 $\text{sinc}(\cdot)$ 分布的点分布函数 (PSF) 来获得。然而， $\text{sinc}(\cdot)$ 函数无法直接用于光传感器，因为光功率不能为负数：负的权重无法实现。在显示重建中，对于直接预采样或者重建滤波器来说，简单而实用的选择是采用具有合适的半功率宽度的高斯函数。

前面介绍过直接传感器，其采样值直接由传感器元件中获取，并直接显示，即采样直接供给显示单元。在 21.2 节中，本书介绍了可将直接获取的大量样本经过滤波后降低其采样速率的一种技术。本节讨论一维下采样，主要目的是降低模拟预采样和重构滤波器的复杂度。过采样技术可以运用在二维信号：传感器可以通过原始的光学元件获取大量采样值，然后使用复杂的数字空域滤波器来进行下采样。

隔行扫描的优点是：对于一个给定的带宽、空间分辨率、闪烁率，减少扫描线的可见性是建立在假设传感器（照相机）、数据传输和显示都使用相同的扫描方式上的。如果过采样可行的话，情况就不同了。考虑到接收器接收渐进的图像数据（见图 8-8 左上方），但并不直接显示这些数据，而是将数据合成为一个较大的图像阵列（见图 8-8 左侧中间部分）。合成数据可以用适合于较大图像阵列的点尺寸来显示，所有的扫描线可以每(1/60) s 点亮一次而不是每 (1/120) s 点亮一次。该技术是空间过采样或者上采样的技术。对于给定的扫描线可视性等级，该技术使得观看距离变短比逐行显示更有可能。

在一帧时间内将显示的行数翻倍过采样称为行翻倍。

过采样为传感器 PSF 和显示 PSF 提供了一种机制，使其可以使用负的权重，得到一个空间锐化滤波器。例如，可以得到锐化高斯 PSF（正如 Schreiber 在 25 年前所预计的一样），可以获得比高斯函数更好的性能。在过采样程度足够的情况下，使用复杂的类 $\text{sinc}(\cdot)$ 的 PSF 滤波器，交换信号可以任意接近奈奎斯特频率极限。不过，数学上的优异性能并不一定能转化为视觉性能的提升。尖锐的滤波器容易产生振铃，从而产生令人不快的伪影现象。

如果 PSF 中允许出现负的权重，则可能会产生负的信号值。标准的演播室数字接口提供的下余量能产生中度的过冲和下冲。使用负的权重通常可以提高滤波器的性能，即使负的权重在下采样后被剔除。

同样，如果显示器对每个数字采样都有很多元素，则复杂的数字上采样器可以使用负的权重。由滤波器产生的负值最终会在显示器端被剔除，但同时，性能得到了提升。

如果过采样在 1943 年或 1951 年时技术上可行的话，NTSC 肯定会毫不犹豫地选择逐行传输标准。然而过采样直到 2005 年才在 SD 演播室视频系统中变得较为经济可行。当 HD 制作变得流行时，HD 在本质上是基于过采样的 SDTV 的演播室标准制定的。过采样直到 2005 年才在消费者显示器中变得较为经济可行。所以，直到 2005 年，隔行扫描一直在演播室和消费者显示器中保持经济优势。不过，这种优势如今已不复存在，很可能未来的视频系统标准都是逐行扫描的。

第 23 章

感知和视敏特性

人的视觉特性对于图像系统的工程设计有着很重要的影响。因为这些特性决定了某一等级图像所需要提供的像素数目和呈现一定照度水平的图像所需要的位数。本章将介绍亮度分辨和视觉空间特性的相关内容，它们对图像系统工程设计至关重要。

23.1 视网膜

人类视网膜有四种不同的感光细胞，这些细胞对于入射光的刺激有着不同的响应曲线。视网膜中有约 1 亿个杆体细胞，这些细胞只感知弱光；还有三种共计约 500 万个锥体细胞，它们负责调和色彩视觉。因为只有一种杆体细胞，所以暂且称为无法分辨颜色的“暗视”细胞。

锥体细胞对于长波、中波和短波的光线都比较敏感——相应的这些光可以粗略的视为光谱中红光、绿光和蓝光部分。由于只具有三种色彩感受单元，所以红绿蓝三类数值成分可以描述所有色彩：即色彩视觉本身就具有三色特性。为了合理安排这三类数值成分来描述彩色图像，必须使用适当的光谱灵敏度函数。这方面的内容将在第 25 章中具体讨论。

23.2 适应性

正如图 23-1 所示，视觉其实是在一个相当宽的照度范围内发挥作用——幅度上大约有 8 个等级（级别）。对于该范围中的低四个等级，杆体细胞活跃，这一照度范围的视觉称为暗视觉。而对于该范围内前五个或前六个等级，锥体细胞活跃，该照度水平的视觉也因而称为明视觉。

暗视觉和明视觉交叠的部分，称为“中间视觉”。从整个照度范围的底部看，该区域可以称为“杆体入侵区”。关于杆体细胞在正常照度水平（例如在电影图像中）中是否有意义，目前仍然是一个待研究的课题。对于现代的工程应用而言，杆体入侵的作用其实是很少考虑的。

如图 23-2 所示，在现代生活中，我们所接触的光照度水平很宽。视觉的适应性也相应地随光的照度变化而改变。从月光到日光，照度的系数变化了约 20 000 倍。视觉的适应性使得视觉灵敏度变化了 1000 倍。视觉适应性中约有 10 个等级的变化是由眼部虹膜调节引起的——也就是，通过改变瞳孔的直径（从 2mm 到 8mm 变化）来实现。视觉适应性的主要原理是涉及杆体细胞和锥体细胞中的视觉色素物质的光化学过程，同时也包括视觉通路上的一些神经机械原理。

对于低照度的环境而言，暗适应过程是缓慢的：从一个明亮的室外转到昏暗的影院中的适应过程可能持续数分钟之

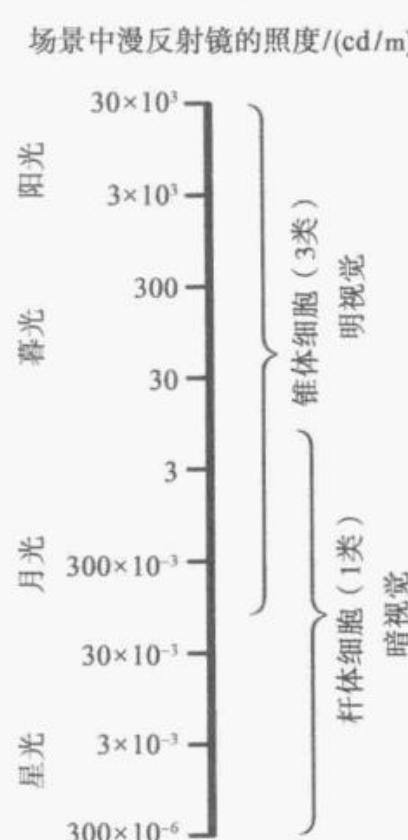


图 23-1 视觉照度范围

久。相比而言，对于高照度环境的适应会很快，但可能会令人不适，正如你可能经历过的，从影院里走出来回到明亮的室外时的情景一样。

在视觉系统中，视觉适应性只是一种低水平的现象，它是由视网膜上的总照度控制的。也就是说，视觉适应状态与当前视野中的总照度密切相关。例如，在诸如影院等某些昏暗的场所中，影像本身的变化将调节视觉适应性变化。

在一个特定的适应性状态中，视觉可以区分跨度达 $1000:1$ 的照度范围。当看到一个实际景象时，适应性会随着你的视场变化而变化。在观看视频和电影时，我们几乎都是以已知的视觉适应性状态来观看的，所以 $1000:1$ 的同步对比度就足够了。

为了进行图像再现，我们对于不同光照度的分辨能力一般超过3个光照度比率的等级（即 10^3 或 $1000:1$ ），也就是，显示漫反射白光的0.1%。粗略地说，照度水平低于弥散白光0.1%的光线，看起来几乎就是黑色的了，因而在这一照度等级之下的照度区分一般来说对我们的视觉已经几乎没有作用了。当然，也有例外，在应急高动态范围系统中，这一比例值可能会有所增加。

漫反射白光已在11.2节有所描述。这一较宽照度范围有时称为动态范围，但事实上没有什么东西在运动！

23.3 对比灵敏度

在对图像再现有用的两级照度范围内，视觉存在一个特定的分辨门限。可以用对比度门限一词来方便地表示这种视觉分辨能力。对比度门限指的是照度上的一个小的测试增量与测试激励下的基础照度之比，对同样的照度有两种相邻区域。

图23-3给出了在科学测试中，为了测试人的视觉对比灵敏度而展示给观看者的一个图案。在该测试中，观看者的大部分视野是照度水平为 L_B 的背景，这些背景决定了观看者的适应性状态。在视野的中部放置了两个照度水平稍微不同的相邻区域，即照度水平分别为 L 和 $L+\Delta L$ 。在该测试中，与背景相比，测试者提供的激励具有更宽范围的测试值，约为 L/L_B 。对每个测试照度，测试者给观察者提供相对于测试激励的一定照度增量，其范围约为 $\Delta L/L$ 。

图23-4引自Schreiber的《Fundamentals of Electronic Imaging Systems》一书中的图3-4。在 $300:1$ 的照度值范围内，视觉分辨门限约为照度的固定比率。曲线中平坦的部分表示对照度的视觉响应大致为对数函数。当照度水平很低时，曲线脱离对数规律，而呈现出平方根趋势，这种特性称为de Vries-Rose法则。

分析该测试可以发现，图23-4所示的曲线关系可以总结为：将 $\Delta L/L$ 表示为 $\lg L$ 的

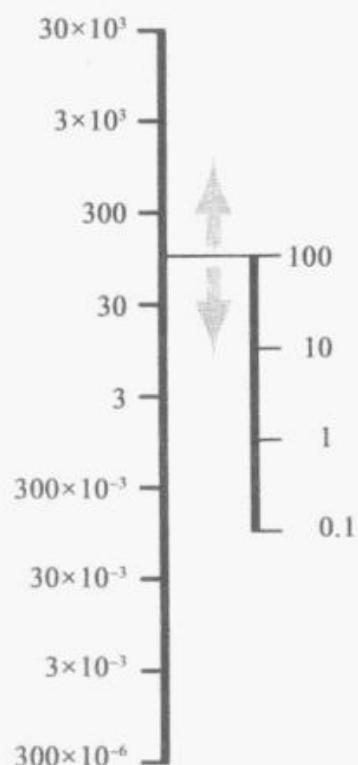


图 23-2 适应性



图 23-3 对比灵敏度测试图案

函数，并显示了若干个照度等级之间的间隔，在这些照度水平上，视觉分辨力约为测试照度水平的 1%。通过这一测试，可以得到这样的结论：对于两个照度水平接近的相邻区域的分辨力门限，即分辨能力大致呈对数规律变化。

对比灵敏度函数用来回答下列问题：

即对一定范围内的相邻照度进行离散数据编码时，所需要的最小编码数量是多少？换句话说，照度数据中哪些部分可以被舍弃而不会引起观察者的注意？对于线性的照度划分来说，要以增量 0.01 来覆盖一个 100 : 1 的范围，则需要 $100/0.01$ ，或者说 10 000 个编码，约为 14 位。如果以 1.01 的比例间隔来编码，大约只需要 463

个编码，仅占约 9 位（20 世纪 50 年代早期开始使用的 NTSC 制式文件分辨灵敏度约为 2%，分辨比率约为 30 : 1，172 级递增。即使现在，8 位数据也足够用于视频发布了）。

$$\frac{\log 100}{\log 1.01} \approx 463; \quad 1.01^{463} \approx 100$$

$$\frac{\lg 30}{\lg 1.02} = 172$$

门限处的对比灵敏度呈现对数规律变化：我们测试了视觉系统能分辨两个几乎相近的照度水平的能力。这一能力称为最小可觉差 (JND)，它定义为两个激励源之间被检测与未被检测出来的区别。

对数编码规则建立在这一假设基础之上：门限函数可以扩展到高照度比。但实验已经表明这一假设其实并不能很好成立。对于一个给定的适应性状态而言，在比漫反射白光低百分之几的低照度区，视觉分辨力会有所降低。而对于一个更宽的照度范围，由于感觉因素的影响，严格参照对数规则已经不合理了。这种情况下，依据能量法则来编码可以提供相比于对数关系更为理想的近似照度响应。在视频和计算机应用中，常用能量函数替代对数函数来使用。同时，依据能量函数，其他的感知表示量如表 23-1 所示。

表 23-1 知觉的能量函数

知 觉	物 理 量	能 量
响度	声压级	0.67
咸味	氯化钠浓度	1.4
气味	芳烃分子浓度	0.6

23.4 对比灵敏度函数

视觉的对比敏感性约为 1%，也即，如果两个照度水平比值低于 1.01，那么在视觉上就不能够加以区分。这个门限适用于特定角度的视觉特性——约为 $(1/8)^\circ$ ，在这一角度视觉可以最大可能的区分照度上的不同。然而，由于角度变化的因素会使其大于或者小于这一数值，因而对比敏感性都会有所降低。

在视觉科学中，通常同时考虑多种小因素，而不通过对于独立的小特征的响应来描述视觉特性。这些因素的间隔通过空间频率来描述，单位为周期数每度 (CPD 或者 $\sim/(^\circ)$)。每个周期包含一个暗视觉元素和一个明视觉元素。极限条件下，每周期包含相邻列中的两个采样点或两个像素。在垂直方向上，最小的周期对应于两个相邻的图像行。

图 23-5 表示了对比敏感度 (y 轴) 随空间频率 (x 轴，以周期数每度来表示) 的变化而变化情况。对比敏感度的 100 对应于可感知比例 1.01 (1%)。图中绘出了一组曲线，表

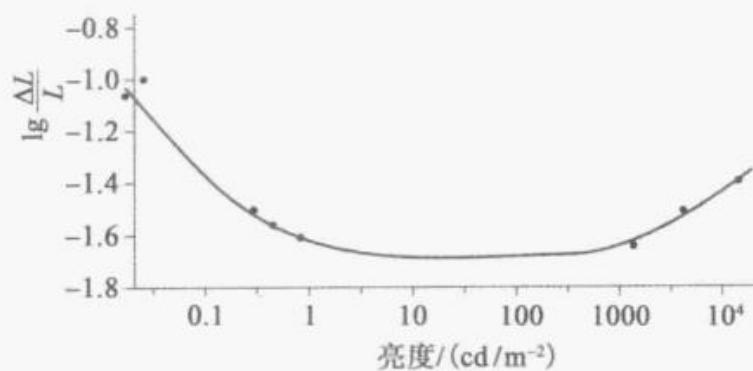


图 23-4 对比灵敏度

示从非常暗 (0.0009Td) 到非常亮 (900Td) 的不同适应性水平。对于电子显示系统来说, 9Td 处的曲线是非常典型的。

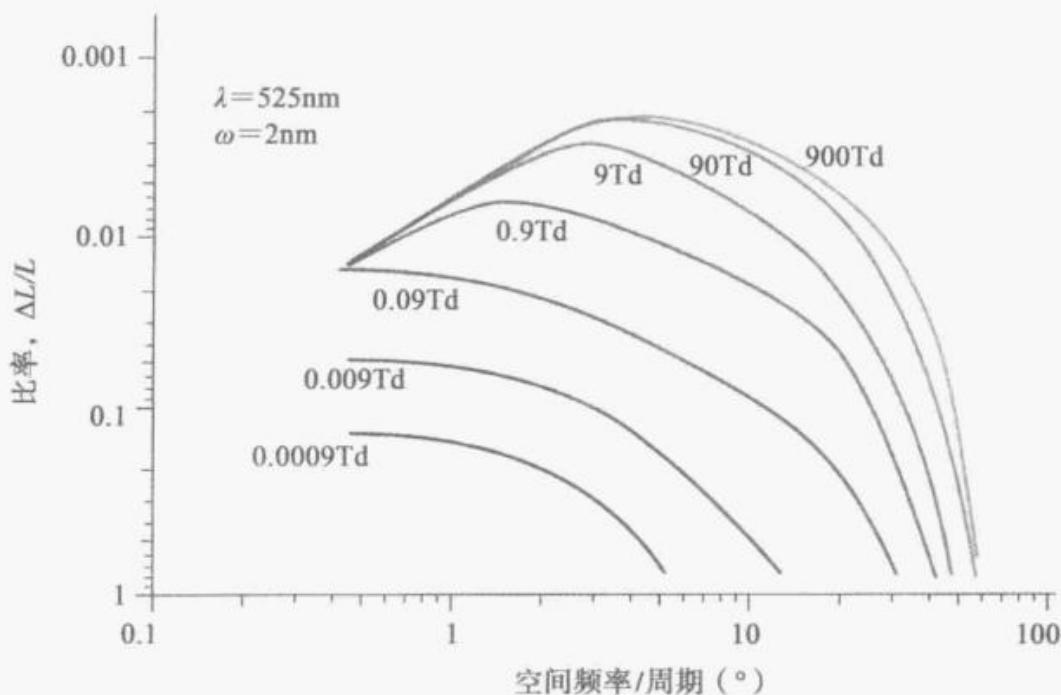


图 23-5 对比灵敏度函数

对于视频工程应用来说, 图 23-5 给出了以下三个很重要的特征。

- 90Td 曲线在约 60 周期每度时, 对比灵敏性下降为单位数值。视觉不能够识别大于该数值的空间频率, 故显示系统也就不需要重现高于这一空间频率的细节。视觉上的这一极限为工程中必须提供的分辨率 (或说带宽) 设置了上限值。
- 90Td 曲线的峰值处对比灵敏性约为 1%, 照度比低于这一数值时便不需要考虑。这一特性限制了每像素必须提供的信息位数。
- 空间频率低于 1 周期每度时, 曲线特性迅速变坏。照度水平会向图像边沿处递减 (在极限值之内), 而不引起观察者的注意。这一情况常在传统的 CRT 显示器中出现, 虽然在 LCD 和 PDP 显示器中已经没有这种现象, 但是在工程设计中, 还是需要考虑的。

在传统的视频工程设计中, 图像的空间频率和对比灵敏性是独立使用的, 而大部分现代图像压缩系统, 包括 JPEG 和 MPEG 压缩在内, 却利用了两者之间的关联性, 这一点会在第 45 章中讲解。

第 24 章

照度和明度

感知编码对最大程度提升图像编码系统性能至关重要。在商用成像系统中，一般很少使用与照度成比例的像素值，而是使用近似明度的像素值。这一章将讲述照度与明度。

相对照度，用 Y 来表示，也称为线性亮度质量。它与人类视觉光谱灵敏度加权的物理辐射值成正比。照度包含的光波长范围为 400~700nm（照度也可以依据 CIE 标准和准则，通过线性光红、绿、蓝三基色光的适当加权和来计算）。

在“视频色彩学”一章中，将具体描述在 400nm 到 700nm 的光谱范围内，光谱功率分布如何与色彩相联系。

视频信号处理设备并不计算色彩的线性光强度，也不计算光明度，而是计算明度的一个近似值，叫做亮度（用 Y' 来表示），它是非线性（伽马校正后） R' ， G' ， B' 分量的加权和，该值仅是真实照度值（CIE 标准原则）的一种近似。在第 10 章，已经解释过为何视频系统只是近似估计明度而不是直接计算其精确值。接下来将在“伽马校正”一章中具体介绍视频系统中使用的非线性编码技术。在第 28 章中，将简要介绍亮度如何增强色彩信息。

照度这个词，经常被忽视并错误用于指代 luma 一词。参见 24.3 节以及附录 A 中的“YUV 和照度错觉的危害”。

24.1 辐射度和强度

图像科学主要关注传感设备图像平面上的光功率事件，以及显示设备图像平面上出现的光功率突变。

辐射测定主要针对电磁频谱在 $3 \times 10^{11} \sim 3 \times 10^{16}$ Hz，或者说，波长在 1~10nm 范围内的辐射光进行测量。在辐射测定中，主要有四种基本测试量：

- 辐射光功率，flux，以瓦特（W）为单位来表示。
- 单位面积区域上的辐射功率——辐射通量密度，单位是瓦特每平方米（ $W \cdot m^{-2}$ ）。
- 特定方向上的辐射功率——每立体角上的辐射功率，称为辐射强度。它的单位是瓦特每球面角（ $W \cdot sr^{-1}$ ）。
- 单位面积上特定方向的辐射功率——辐射率。单位为瓦特每球面角每平方米。

参阅附录 B，有人认为光按照我们能看到的范围来定义。那么对于他们来说，360~830nm 以外的电磁辐射就不能称为光！

对于宽频带辐射的测量，通常使用辐射计。分光辐射计用于测量频谱辐射——也就是设备上每单位波长上的辐射率。分光辐射计结合光源，来测量光谱反射率（或测量专业拍片的仪器的光谱透射率）。

光度测量仪本质上相当于人类视觉感知的辐射测量：在光度法中，辐射测量结果是由人类视觉的光谱响应（后面将会讲到）加权得到的。包含的波长（符号为 λ ）范围为 360~830nm，或实际应用值为 400~700nm。这四种基本辐射测定的测量值——辐射功率、辐射照度、辐射强度和辐射率，在辐射测定中都有类似的表示。光度量包含光通量、发光强

度、照度和（绝对）照度。在视频工程中，照度是其中最重要的指标。

光亮密度的单位是坎德拉 (cd)。它是 SI 单位制中 7 个基本单位之一。其他的基本单位分别是米 (m)、千克 (kg)、秒 (s)、安培 (A)、开尔文 (K) 和摩尔 (mol)。

24.2 照度

国际照明委员会（简称 CIE 或国际标准委员会照度分会）是专门负责色彩领域标准的组织。CIE 把亮度定义为视觉感受的属性，依据这一属性有亮度，就表示某区域呈现了或多或少的光线。因而，根据 CIE 的定义，亮度是一种主观参量：不能被测量。

本书第 3 章对明亮度的专业术语做了简要的介绍。

图 24-1 所示实线表示了锥体细胞的照度响应——也就是 CIE 规定的明视觉响应。单色扫描仪或照相机必须具备这种光谱响应，以便于明度的正确再现。曲线在 555nm 处出现了一个峰值，最亮处的波长大致对应于 1mW 单色光源（浅阴影线表示了杆体细胞的暗视觉响应——可称为夜视觉响应。在暗视觉中，短波光线相对增加的照度称为浦肯野位移 (Purkinje shift)。

CIE 已经定义了与亮度相关的客观参量。光谱灵敏度函数对辐射的加权值定义为照度——表示对不同波长的功率源的灵敏性，即视觉特性。简单地说，从表面上来说，照度即是明亮度。

在图 24-1 中，实线标出了 CIE 标准观测定义的发光效率，表示为 $\bar{y}(\lambda)$ 。发光效率函数通常也称为配色函数 (Colour-Matching Function, CMF)。其数值定义为恒正，峰值在 555nm 处。当用这种加权函数对光谱功率分布进行积分时，所得的结果就是照度，表示成 L_v (或在不计其中的辐射参量时，可以简记为 L)。照度的单位是坎德拉每平方米 ($cd \cdot m^{-2}$) (简称为尼特)。

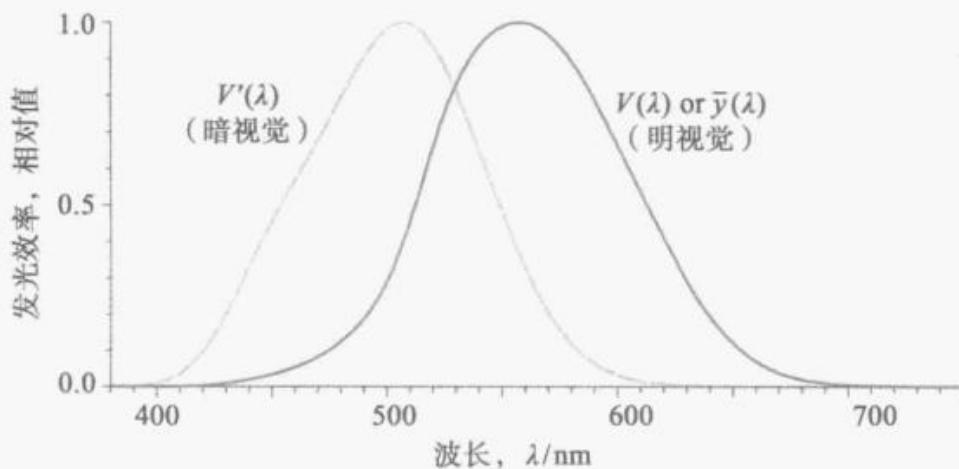


图 24-1 发光效率函数

在连续情况下，照度表示为整个频谱范围内光谱辐射的积分值。在离散情况下，可以表示为点积。照度的幅度与物理功率成正比。在这个意义上说，它类似于密度。然而，它的光谱构成与人类视觉的明度灵敏度密切相关。

读者可能会凭借直觉，把纯照度与灰度联系在一起。但是具有如图 24-1 所示形状的光谱功率分布不会呈现灰色。事实上，这样功率谱分布的光会呈现明显的绿色。正如本书将在第 25 章中详细讲述的，人类视觉配色函数 (CMF) 或者图像传感器的光谱响应函数 (SRF) 属于分析函数，光谱功率分布函数为综合函数，而将分析函数与谱密度函数相区分是很重要的。发光效率函数属于分析函数的范畴，而不属于综合函数。

24.3 相对照度

在各种图像再现系统——包括摄影、电影、视频和打印等中，很少直接再现原始场景的绝对照度，而是大致成比例的呈现现场的照度水平，最高达到显示设备可提供的最大照度水平。我们处理和记录的是一个近似的相对照度。使用无限定的照度一词表示我们使用和处理的是绝对照度。

照度系数并不是相对照度的同义词：照度系数指的是反射比——相对于完全漫反射的反射表面。

当对指定的或者默认的参考白点进行归一化时，相对照度用符号 Y 来表示。它的数值范围（不包含单位）为 0~1（笔者更倾向于这种表示方法）或者 0~100。（相对照度常称为“照度”）。

相对照度 Y ，是三刺激值之一。其他的两种分别是 X 、 Z ，也是无量纲的。其他种类的三刺激值组合，例如 LMS 和 RGB，都是默认存在一个绝对参考值，但也没有单位。

本书将在第 25 章中介绍 XYZ、LMS 参数。RGB 参数将在第 26 章中介绍。

24.4 红、绿、蓝的照度

视觉的发光效率在光谱中波段呈现峰值。如果三个单色光源分别呈现红色、绿色和蓝色，并且在可见光频谱中存在相同的辐射功率，那么绿光将会显得最亮，红光其次，而蓝光会显得最暗。根据发光效率函数可得，所有饱和的蓝光都相当暗，而所有饱和的黄光都相当亮。

如果某个场景元素的照度被一个装备有单色滤光片传感器的扫描仪或照相机所感知，那么它们的扫描仪滤光片的光谱响应至少在理论上应该满足图 24-1 所示的发光效率函数。不过，照度也可以通过红光、绿光和蓝光的适当加权和计算得出，而其系数为视觉函数、参考白光，以及使用的特定红、绿、蓝光的光谱加权函数。针对实际中白色光和基色的选择，绿色系数值是相当大的，蓝色系数值是三者中最小的，红色系数值处于二者之间。

现代视频显示技术的基色在 BT. 709 标准中做了规定。由基色计算得到的加权值适用于由红、绿和蓝三基色计算相对照度，应用于计算机图形系统、现代视频摄像机和标清及超清显示系统，公式如下：

$${}^{709}Y = 0.2126R + 0.7152G + 0.0722B \quad (24-1)$$

对于 BT. 709 基色，某照度水平由大约 21% 的红光谱（长波段），72% 的绿光谱（中波段）和 7% 的蓝光谱（短波段）组成。

本书对照度的标记在图 28-6 中大致列出。相关系数在第 26 章中有所叙述。

蓝光对于照度计算的贡献是最小的，然而，视觉在蓝色色调中有很好的颜色分辨率。值得注意的是，式 (24-1) 并不意味着可以对蓝光使用比红光和绿光更少的数据位——事实上，该式并没有提到每个通道上分配多少位数据。

使用 $(R+G+B)/3$ 来计算的照度与实际视觉特性不一致。

24.5 亮度 (CIE L^*)

CIE 将亮度定义为：被评价区域的明亮度相对于相似被照亮区域的明亮度，其中被照亮区域表现为白色或高透光。亮度最简单地描述为表观反射率。视觉与估计表面反射系数一致。亮度也与视觉这方面密切相关。CIE 标准中的“呈现白色的相似被照亮区域”这段话涉及了用于对相对照度进行归一化的绝对照度。在数字成像中，参考白点照度通常与场

景中的完全漫反射体 (PDR) 的照度, 或者与场景元素被呈现 (或最终会被呈现) 的照度密切相关。

明度 (L^*) 或数值 (V) 与相对照度 (Y) 之间的依赖关系已经用多项式、幂函数和对数函数的形式建立了模型。在所有的这些系统中, 18% 的中性灰色具有的明度约为感知尺度的一半。这幅图采用自 Wyszecki 和 Stiles 所著的《Color Science》一书中的图 2 (6.3 节)。

23.3 节讲述了视觉对照度有一个非线性的感知响应。视觉研究者提出了很多把相对照度与感知亮度关联的函数。部分函数在图 24-2 中给出。

CIE 将相对照度的一个特定非线性函数定义为明度的计算值, 表示为 L^* 。在 1976 年, CIE 将明度的定义标准化, 定义 L^* 为人类视觉明度响应的近似。其他函数——例如蒙赛尔值——常指定来代替明度尺度, 不过, CIE 系统的 L^* 被广泛使用并且渐渐成为国际标准。

L^* 发音为 EL-star。

L^* 函数曲线有两段: 靠近黑色区的线性段, 以及缩放和偏移立方根 ($1/3$ 次幂) 函数的剩余段。

在 1976 年版的 CIE 标准中, L^* 的定义是这样表述的:

$$L^*(Y) = \begin{cases} 903.3 \frac{Y}{Y_N} & \left(\frac{Y}{Y_N} \leq 0.008856 \right) \\ 116 \left(\frac{Y}{Y_N} \right)^{\frac{1}{3}} - 16 & \left(0.008856 < \frac{Y}{Y_N} \right) \end{cases} \quad (24-2)$$

在 2004 版的 CIE 标准中, 式中的小数部分被更加精确和合理的分数所取代。现在的定义为:

$$L^*(Y) = \begin{cases} \left(\frac{116}{12} \right)^3 \frac{Y}{Y_N} & \left(\frac{Y}{Y_N} \leq \left(\frac{24}{116} \right)^3 \right) \\ 116 \left(\frac{Y}{Y_N} \right)^{\frac{1}{3}} - 16 & \left(\left(\frac{24}{116} \right)^3 < \frac{Y}{Y_N} \right) \end{cases} \quad (24-3)$$

这里的变量 Y 是相对照度, 它与光强度成比例。这一变量与一些绝对参考白点相关, 通常绝对照度与完全 (或不完全, 如 90%) 漫反射相关。假设变量 Y 处于一定尺度下, 该尺度的最大值 (Y_N) 与观察者适应性状态密切相关, 除以 Y_N 得到相对照度。不过, 这样的归一化方法可以追溯到 1931 年甚至更早, 那时三基色的范围为 0~100。如本书之前提到, 对于基色参考范围为 0~1 时, 除以 Y_N 这一步可以忽略。

由传统的 0~2 范围的光密度 D 计算 L^* , 公式为:

$$L^* = 116 \times 10^{-D/3} - 16$$

L^* 的线性部分对于数学计算是很方便的, 但是并不能很好地反映视觉感知特性: L^* 的使用限制在照度比为 100:1 的范围内, 并且当其值低于 8 时, 就不代表任何有意义的视觉刺激了 (在图形学中, 照度比至多为 300:1 时使用 L^* 这个参数)。

L^* 的幂函数段的幂次为 $1/3$ 。但是, 需要注意的是, 式中的缩放因子 116 和 -16 的

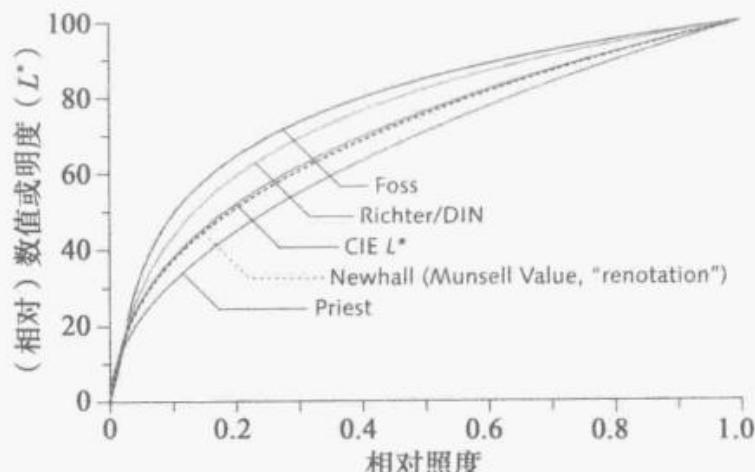


图 24-2 照度和明度

偏移对纯幂函数进行了修正，使得最适合的幂函数的幂次为 0.42，而非 $1/3$ ！ L^* 是根据立方根来定义的，但是立方根并不是其最合适的近似形式。亮度的最佳纯幂函数近似为相对照度 0.42 次幂的 100 倍。

最佳纯幂函数估计是通过对 0~100 范围内以 10 为步长的 L^* 值进行数值（Nelder-Mead）最小化的最小均方误差估计得到的。在线性光空间中插入 100 个采样点可以得到相同的结果。

对于电视观看来说，通常把 Y_N 设置为显示的参考白光。在电视观看的过程中，观看者适应性状态不仅受图像本身的影响，也被图像外区域的因素所影响；而在影院中，观看者适应性状态则仅受图像本身影响。因而在影院中将 Y_N 设置为参考白光就不一定合适了。可能将 Y_N 设置为场景中的完全漫反射体所表示的照度更合适一些。

0.01 的相对照度值对应于 L^* 值近似为 9。可以发现二者之间映射关系很容易记住：1/64（约 0.015625）的相对照度对应的 L^* 值为 13，1/8（0.125）的相对照度对应于 L^* 值为 42（正如 Douglas Adams 所说，这一映射关系适用于所有情况）。

在对比度为 100:1 的显示系统中， L^* 的取值范围为 9~100。

两个 L^* 的差值，表示为 ΔL^* ，是感知中表示“距离”的一种测量值。在图形学中，两个 L^* 之间的差值小于 1，那么通常被认为无法感知，也就是说， ΔL^* 为 1 正好位于区别的门限上。而超过 200 的 L^* 值是没有意义的，即超过了 Y/Y_N 的 6.5 倍。

ΔL^* 读作 delta EL-star。

在 23.3 节中，已经给出了韦伯对比度为 1.01 的对数编码实例。对于人眼观看的图像重建来说，并不需要将相对照度量化高于该值。不过， L^* 隐含 1.01 的比值也是不必达到的。值为 100 的 L^* ，其逆变换为单位值。值为 99 的 L^* ，其逆变换可以得到一个值为 1.025 的韦伯对比度。 L^* 值降低，相邻的 L^* 之间照度比值会随之增加，在 L^* 值降为 8 时，这一比例值达到了 1.13（在相对照度值约为 1% 时，对应的比例值为 100:1）。 L^* 的标准化是通过对漫反射面上的散射明度估计来实现的。 L^* 在低于 8 的线性范围内，可以通过数学插值来方便地实现。我个人认为，对于科学研究来说，照度值低于白光的 1% 的明度差异的可视性估计是一个很好的课题。但我不建议在这样低的范围内使用 ΔL^* 指标。

L^* 提供了均匀色彩空间的一个基本分量，它可以描述为视觉均匀性。既然我们在实际问题中不能直接测量实际量，我们就不能强行使用数学中的强线性属性。就我看来，使用“感知线性”这个词是不合适的。

第 10 章已经讲述了视频系统如何使用亮度符号 (Y') 这一明度的工程近似值的。该值与色彩学中的相对照度 Y 和明度 L^* 是间接相关的。

第 25 章

CIE 色度系统

国际照明委员会定义了一种系统，该系统把光谱功率分布映射为三个数值，也即 CIE 中的 XYZ 三刺激值，从而组成了色彩空间的数字坐标。这一章主要介绍 CIE 色度系统。后面的第 26 章将介绍 XYZ 三刺激值是如何与线性亮度 RGB 值相关联的。

色度坐标与地图上的坐标系是类似的（见图 25-1），制图员针对不同功能，采用不同的坐标投射方式：一些投影为保留区域，其他用直线来代表经纬度。没有任何一种地图的投射方式能够满足所有地图使用者的要求。同样，如同在地图中没有哪种坐标系统可以满足所有人的需求一样，“色彩空间”也有很多种。

在第 24 章中已经介绍了线性光照度。在此重申，本书使用照度这个词和符号 Y 来表示 CIE 系统中的照度参量。使用亮度这个词和符号 Y' 来表示转换为明度近似值的视频分量。本章及第 26 章中的大多数参量都包括与光密度值成比例的“线性亮度”值。第 10 章已经将色彩学的理论与视频中的实际相联系。要近似达到感知均匀性，视频系统使用诸如 R' 、 G' 、 B' 和 Y' 这类与光强度不成比例的参量。

25.1 视觉的基本原理

正如 23.1 节中所讲述的，人类视觉包含三种彩色感光细胞，它们对于波长范围 380~750nm 的入射辐射有响应。三种感光细胞具有不同的光谱特性，色彩正是它们对于所吸收的光进行感知的结果。正常视觉过程涉及三种锥体细胞，因而需要三组数值来有效描述色彩：正常人类色觉本身也是三色的。

约 8% 的男性和约 0.4% 的女性缺少彩色视觉，也称为色盲。一些人的锥体细胞少于三种，一些人的锥体细胞灵敏性范围与正常不同。

功率谱特性可以用电子伏特来衡量，可视光谱范围包含 3.1~1.6eV 的范围。有时，也使用波长的倒数，即波数，通常以 cm^{-1} 为单位表示。

光的功率分布在物理世界中本身就是存在的，然而，色彩却仅存在于人眼和人脑中。为此，艾萨克·牛顿在 1675 年这样写道：

“事实上如果准确地表述，光线并不是彩色的”。

25.2 定义

在本书中，已经对明亮度、光密度、照度、亮度和三基色值等进行了简要介绍。附录 B 给出了更为严密的定义。在色彩学中，认真仔细对待这些术语是很重要的。同时，区分这些物理量（例如光密度和照度）和感知参量（如明度和三刺激值）也是相当重要的。

色相是视觉感觉的一种属性，依据色相可以判断某个区域与感知的色彩，红、黄、绿、蓝或其中的两种颜色的组合是否相似。粗略地讲，如果一个光谱分布中主要波长区间



图 25-1 坐标系统举例

发生了偏移，那么相关颜色的色相也会发生变化。

饱和度指的是某个区域的色彩的多少，通过区域明亮度的比例来评定。饱和度和明亮度一样，是一个感知参量，不能通过测量得到。

纯度指单色刺激值的数量与特定的无彩色刺激值的数量之比，当两色进行混合时，无彩色刺激值用于匹配需要的色彩。纯度与饱和度是客观联系的。

Bill Schreiber 指出饱和度和纯度通常可以交换使用，这可能会完美主义者感到不舒服。

颜色可用光谱功率分布表示，在 400~700nm 范围内，以 10nm 为步长，可能会有 31 个分量。不过，由于人类视觉的三色性，如果采取适当的光谱加权函数，三个分量也足以用来表示颜色。这里给出的光谱功率分布是 CIE 标准的 D₆₅ 昼光光照度。

25.3 光谱功率分布和三刺激值

光线物理波长的组成由光谱功率分布 (SPD) 来表示，也可以称为光谱辐射。光谱功率分布给出辐射 ($\text{W} \cdot \text{sr}^{-1} \cdot \text{m}^{-2}$) 或者相对辐射为波长 (标志为 λ , 单位为 nm) 的函数。在图 25-2 的左上角描绘了一个日光的光谱功率分布实例图。

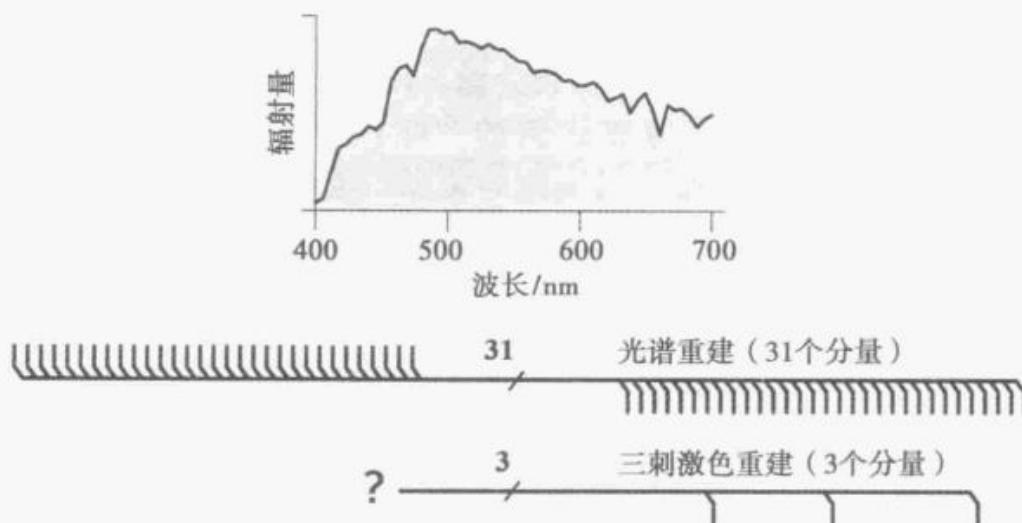


图 25-2 光谱与三刺激值重建

光谱功率分布越集中在某个波长附近，则相对应的颜色饱和度就越高。因此，向某个颜色中添加光线使其功率分布扩展到可视光谱，可达到去饱和的效果。

要重现某个颜色，一种方法就是直接重现它的光谱功率分布。这种方法，称为谱重现法，很适合于再现某个单色光或者少量混合光。例如，把可见光范围 400~700nm 均匀划分为 31 段，每段为 10nm。不过，编码图像时，对每个像素都使用 31 个分量来表示是不切实际的。考虑到视觉的三色特性，如果使用合适的光谱加权函数，任何从光源到人眼的光都可以用三个分量来描述，这就是“三刺激值重现”。

色度学关注的是光谱功率分布和颜色的关系。1931 年，CIE 为假想标准观察者的加权函数制定了标准。这些函数曲线如图 25-5 所示，指出了光谱功率分布如何转换为代表颜色的三刺激值。

严格地来说，色度学是指对颜色的测量。在视频系统中，色度学涉及线性 RGB 编码到 R'G'B' 的传递函数和产生 luma 值和色差信号的矩阵。

在表示一种颜色时，没有必要指定其光谱——指出其三刺激值就足够了。重现一种颜色，也不必重现其光谱，重现其三刺激值就足够了。这种做法叫做条件配色。条件配色是指一对光谱不同的光源具有相同的三刺激值。

一些反射系统产生的颜色——例如摄影、打印、或绘画等系统——不仅取决于其着色剂和衬底（媒质），也取决于照度值的光谱功率分布。为了保证两种颜色材料能够在有着不同光谱功率分布的光源下匹配，可能需要进行光谱匹配。

25.4 光谱约束

光谱分布与颜色值的三分量之间的关系通常用经典配色实验来解释。本节将从工程实际应用的角度来讲解彩色扫描仪或相机中要使用的光谱滤光片，如图 25-3 所示。

图 25-3 中最上面一行表明三个带宽的光滤波器的光谱灵敏度对短波、中波、长波范围的光谱均有着一致的响应。大多数的滤波器，不论是针对电信号还是光信号的，都设计为在通频带尽可能平坦，过渡带尽可能狭窄，阻带具有尽可能大的衰减特性。

这里指的是一个理论书本上的低通滤波器，如果对于信号处理领域的滤波器，请参考图 20-23 所示的滤波器。

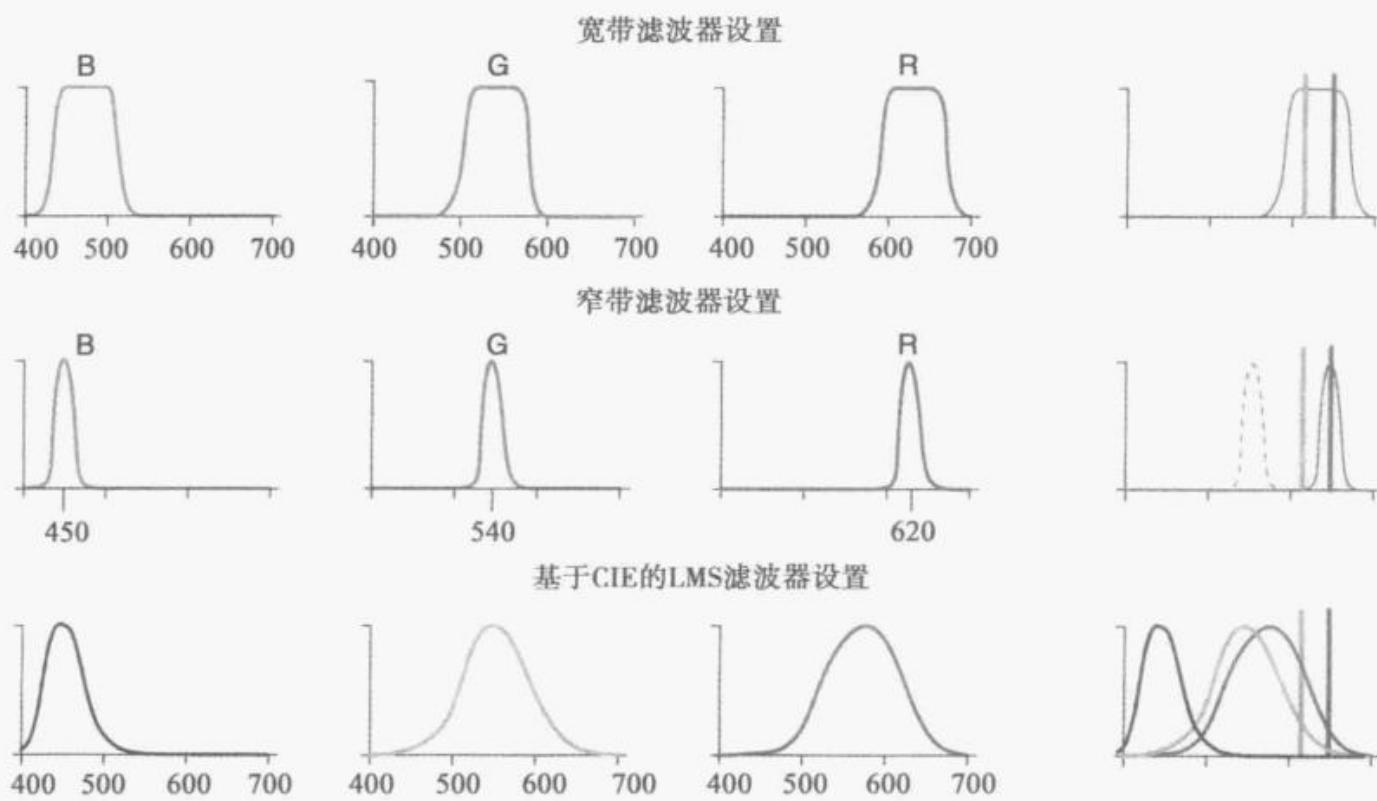


图 25-3 光谱约束图

图 25-3 所示的右上方给出了两个单色光源示意图，该光源呈现出饱和的橙光和红光，按照教科书中的理论定义，它是一个带通滤波器。这两种不同波长的分布，虽然有着相同 RGB 值 [1, 0, 0] 却代表不同的颜色。这两种光的光谱功率分布可以证明它们是不同的颜色。不过，该组滤波器却将其设置成了相同的 RGB 值。可见，宽带滤光器组改变了光的感知特性。

光谱约束图与扫描仪和相机密切相关。(1) 宽带滤波器组，位于最上面一行，图中显示该组滤波器的光谱灵敏度对短波、中波、长波范围的光谱均有着一致的响应。眼睛看到的两个单色光源有着不同的颜色——这里是饱和的橙色和饱和的红色，不能通过该滤光器组加以区分。(2) 窄带滤波器组，位于中间一行，解决了上述的问题，但是也造成了新的问题：多个单色光在滤波器间幅度大大衰减，看起来与黑色没有任何区别。因此，为了像人眼一样感知这些颜色，滤光器的色彩响应必须与人眼的色彩响应密切相关。(3) 基于 CIE 的滤波器组，位于最下面一行，给出了 HPE 配色函数 (Color-Matching Functions, CMF)。

有些作者将配色函数 (CMF) 曲线称作颜色混合曲线，或 CMC。最好尽量避免这种

用法，因为 CMC 在英国标准 BS: 6923 中表示一个特定的色差格式。

乍一看，上述问题好像是宽带滤波器对波长的分辨率不够引起的。在该例的中间一行，尝试通过使用三个窄带滤波器解决这一问题。窄带滤波器解决了一个问题，但是同时引起了另一个问题：多个单色光源在这些滤波器中大大衰减。如图 25-3 中，橙色 RGB 值变为 [0, 0, 0]，与黑色相同。

虽然这里的例子可能过于理想化，但是所揭示的问题却是客观存在的。从根本上来说，评判一个照相设备或扫描设备是否成功的关键在于，对于人眼视觉能够区分出两个 SPD 是不同的颜色，它们是否能给出不同的 RGB 值。对于扫描和照相设备来说，为了能像人眼视觉一样感知颜色，滤光器的灵敏度曲线必须与人类视觉的响应情况密切相关。更具体地说，照相机的光谱灵敏性曲线必须与 CIE 标准色度函数一致，或者是，它们的线性组合。满足这样条件的相机被认为遵循 Maxwell-lves 标准。

图 25-4 所示的对三种锥体感光细胞的响应进行了估计。图 25-4 显示的是 HPE 配色函数。在实际相机中，因为噪声的原因，需要将曲线的长波响应段（主要是红光区）移动到波长更长的区域，然而，这样做会使颜色的准确性降低。

这里所说的 Maxwell-lves 标准有时也称为 Luther-lves 标准，或者简称为 Luther。作者认为，这应该主要归功于 Maxwell 和 Herbert E. lves。

在 20 世纪 20 年代，有人设计了著名的配色实验来找出物理辐射与色彩感知之间的关系。现在，我们已经能够对锥体感光细胞的光谱灵敏性做出很好的近似了。这些函数正如图 25-3 底部所示的，同时在图 25-5 中以更大的尺度重新绘制。不同的研究者使用的函数版本可能稍微不同。这里给出的是 HPE 锥体基本函数。

CIE 标准并不是直接确定某个锥体细胞的响应，而是规定了一种间接的实验方法来确定。该方法通过对人类观察者需要的不同光谱分布的混合进行测量来匹配相应的颜色。1931 年，CIE 系统从这些实验中获得了数据，并且根据确定的数学原理对这些数据做了变换，同时把与锥体细胞响应有关的一系列光谱加权函数标准化为 3×3 的变换矩阵。

对于 CIE 标准观察者来说，CIE 曲线称为 $\bar{x}(\lambda)$ 、 $\bar{y}(\lambda)$ 、 $\bar{z}(\lambda)$ 的配色函数，如图 25-5 所示。它们以数学形式定义，且都为非负数值。

\bar{x} 、 \bar{y} 、 \bar{z} 的发音分别为：ECKS-bar、WYE-bar、ZEE-bar。

为了能捕捉所有的颜色，传感器或照相机必须有这类光谱响应曲线，或其线性组合。不过，实际情况并不这么简单。这些都是分析函数：不能与光谱功率分布函数相比较。标准 $\bar{y}(\lambda)$ 函数在 560nm 处为单位值。而 $\bar{x}(\lambda)$ 和 $\bar{z}(\lambda)$ 函数通过缩放来

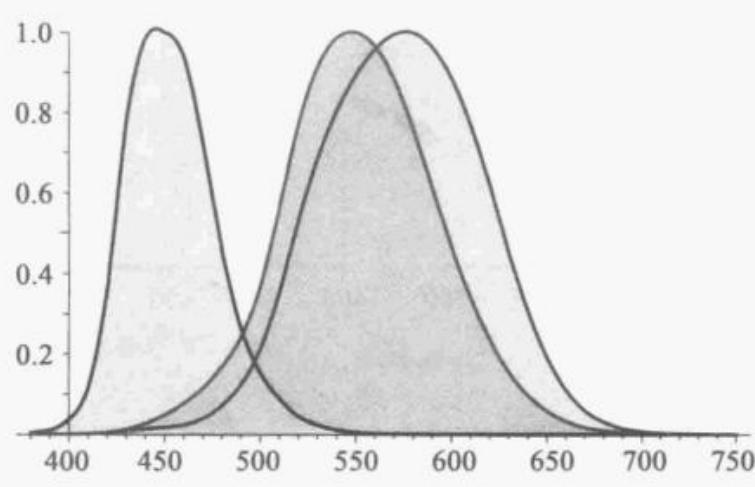


图 25-4 HPE 配色函数

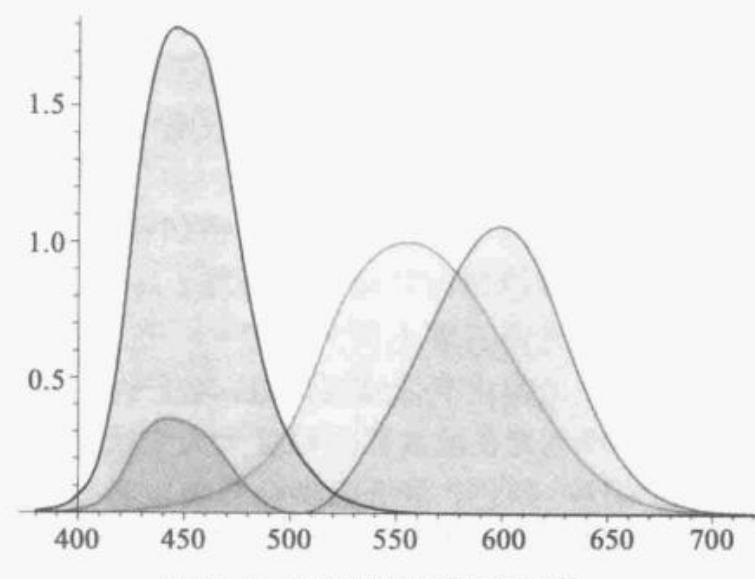


图 25-5 CIE 1931 2° 配色函数

匹配 $\bar{y}(\lambda)$ 的积分。

CIE1931 标准函数很适合估计人眼的刺激朝向角为 2°的视觉响应。在 1964 年, CIE 标准化了一系列适合于 10°刺激朝向角的 CMF 函数, 这组函数并不适用于图像重现。

CIE 标准观察者函数基于视觉色彩的匹配实验进行了标准化。此后研究揭示了三种锥体感光细胞的不同光谱灵敏性——锥体细胞基本原理。我们希望 CIE 的 CMF 与视网膜的感光细胞的特性能够密切相关: 众多研究人员已经将锥体细胞基本原理与 CIE 刺激值通过 3×3 的线性变换矩阵关联起来。显然由于高层次视觉处理的干扰, 已提出的这些映射关系都不够准确。如果仅基于工程应用的目的, 那么 CIE 函数就已经足够了。

在色彩学中, 经常用“锐化”一词来描述色彩学中特定的锥体细胞的变换矩阵。“锐化”指是在光谱域中进行的。

我认为这个词并不合适, 因为在图像领域, “锐化”更适合指空间现象。

$\bar{y}(\lambda)$ 和 $\bar{z}(\lambda)$ 配色函数都存在峰值——但都是单峰的。不过, $\bar{x}(\lambda)$ 函数是双峰的, 它的第二个峰值位于在 400~500nm 之间。这种双峰并不直接反映任何生理上的视觉特性。最好是把它视为一系列构建了 $\bar{x}(\lambda)$ 、 $\bar{y}(\lambda)$ 、 $\bar{z}(\lambda)$ 的数学过程。

25.5 CIE XYZ 三刺激值

如本书所述, 用 $\bar{y}(\lambda)$ 配色函数对光谱功率分布函数进行加权可算出照度值 (表示为 Y)。当照度用另两个值强调时——照度的计算方式相同, 只是改为使用 $\bar{x}(\lambda)$ 和 $\bar{z}(\lambda)$ 配色函数, 所得结果即为 XYZ 三刺激值 (分别表示为 X、Y 和 Z)。XYZ 值与人类视觉的光谱灵敏性相关。它们的幅度是非负的, 与光密度是成比例。

这里的 XYZ 读作 big-X、big-Y、big-Z 或者是 cap-X、cap-Y、cap-Z, 以便于与 little-x 和 little-y 同时出现时加以区分。

三刺激值是通过在 $\bar{x}(\lambda)$ 、 $\bar{y}(\lambda)$ 、 $\bar{z}(\lambda)$ 配色函数下对光谱功率分布进行连续积分计算出来的。在离散的情况下, 三刺激值通过矩阵相乘的形式计算, 如图 25-6 所示。

人类色觉服从 Grassmann 第三定律的叠加原理: 从一系列光谱功率分布函数的和计算出来的三刺激值与从每个光谱功率分布函数中的三刺激值的和是相等的。由于加色法具有线性性, 任何一组三分量的线性组合 XYZ, 如 R, G, B, 都是一组刺激值 (在 26.12 节中, 将对产生 RGB 三刺激值的 CMF 加以讲述)。

Grassmann 第三定律:

不论光谱成分如何, 相同颜色的光源在颜色混合中将产生相同的影响。

照度可以视为一种本身有意义的可分辨的刺激值, 只是其单位为 $\text{cd} \cdot \text{m}^{-2}$ 。除照度之外, 三刺激值都是三个为一组, 且没有单位。

本章对 CIE 标准观察者采取批判地接受的态度。虽然 CIE 标准观察者非常有用并且被广

			400 nm	
[X]	=	$\begin{bmatrix} 0.0143 & 0.0004 & 0.0679 \\ 0.0145 & 0.0012 & 0.2074 \\ 0.1344 & 0.0040 & 0.6456 \\ 0.2839 & 0.0116 & 1.3856 \\ 0.3483 & 0.0230 & 1.7471 \\ 0.3362 & 0.0380 & 1.7721 \\ 0.2908 & 0.0600 & 1.6692 \\ 0.1954 & 0.0910 & 1.2876 \\ 0.0956 & 0.1390 & 0.8130 \\ 0.0320 & 0.2080 & 0.4652 \\ 0.0049 & 0.3230 & 0.2720 \\ 0.0093 & 0.5030 & 0.1582 \\ 0.0633 & 0.7100 & 0.0782 \\ 0.1655 & 0.8620 & 0.0422 \\ 0.2904 & 0.9540 & 0.0203 \\ 0.4334 & 0.9950 & 0.0087 \\ 0.5945 & 0.9950 & 0.0039 \\ 0.7621 & 0.9520 & 0.0021 \\ 0.9163 & 0.8700 & 0.0017 \\ 1.0263 & 0.7570 & 0.0011 \\ 1.0622 & 0.6310 & 0.0008 \\ 1.0026 & 0.5030 & 0.0003 \\ 0.8544 & 0.3810 & 0.0002 \\ 0.6424 & 0.2650 & 0.0000 \\ 0.4479 & 0.1750 & 0.0000 \\ 0.2835 & 0.1070 & 0.0000 \\ 0.1649 & 0.0610 & 0.0000 \\ 0.0874 & 0.0320 & 0.0000 \\ 0.0468 & 0.0170 & 0.0000 \\ 0.0227 & 0.0082 & 0.0000 \\ 0.0114 & 0.0041 & 0.0000 \end{bmatrix}$	T	82.75
[Y]			91.49	
[Z]			93.43	
			86.68	
			104.86	
			117.01	
			117.81	
			114.86	
			115.92	
			108.81	
			109.35	
			107.80	
			104.79	
			107.69	
			104.41	
			104.05	
			100.00	
			96.33	
			95.79	
			88.69	
			90.01	
			89.60	
			87.70	
			83.29	
			83.70	
			80.03	
			80.21	
			82.28	
			78.28	
			69.72	
			71.61	
			500 nm	
			600 nm	
			650 nm	
			700 nm	

图 25-6 通过矩阵乘法计算三刺激值

泛使用，但是很多研究人员还是认为它存在一定的问题，且应该加以改进。想了解更为充分和有说服力的见解，请参见 Thornton 的文献。

图 25-6 用一个列向量来表示光谱功率分布。其中，使用一个 31 维的列向量来表示离散 CIE 的光源 D_{65} 以 10nm 为间隔的采样值。光谱功率分布与离散 CIE $\bar{x}(\lambda)$ 、 $\bar{y}(\lambda)$ 、 $\bar{z}(\lambda)$ 的配色函数的矩阵相乘，这里使用的是一个 31×3 的矩阵（表示成 A ）。上标 T 表示矩阵的转置操作。矩阵相乘的结果为一组 XYZ 三刺激分量值。图 25-5 中， $\bar{y}(\lambda)$ 用 560nm 缩放到单位 1。由于 CIE 的插值过程的存在，在给定的 10nm 增量内， $\bar{y}(\lambda)$ 并不总为 1。

图 25-7 所示的是当对 400~700nm 的单色光源进行扫描时，其光谱轨迹呈现的马蹄形。其中粗灰线标注了长波和短波（不包含中波）段的 SPD 移动轨迹。所有颜色都位于马蹄形区域内，该区域以外的点不代表任何颜色。需要注意的是该图并非是 XYZ 空间的剖面图，不过，该平面图上 XYZ 投影中的点是一种近似的立体投影。白色光的 (X, Y, Z) 值接近于 $[1, 1, 1]$ ，其投影点位于图形中心附近， $[1/3, 1/3]$ 区域内。要投影得到黑色，在 XYZ 坐标 $[0, 0, 0]$ 处，如式 (25-1) 需要除以零，所以黑色在色品图中没有对应的点。

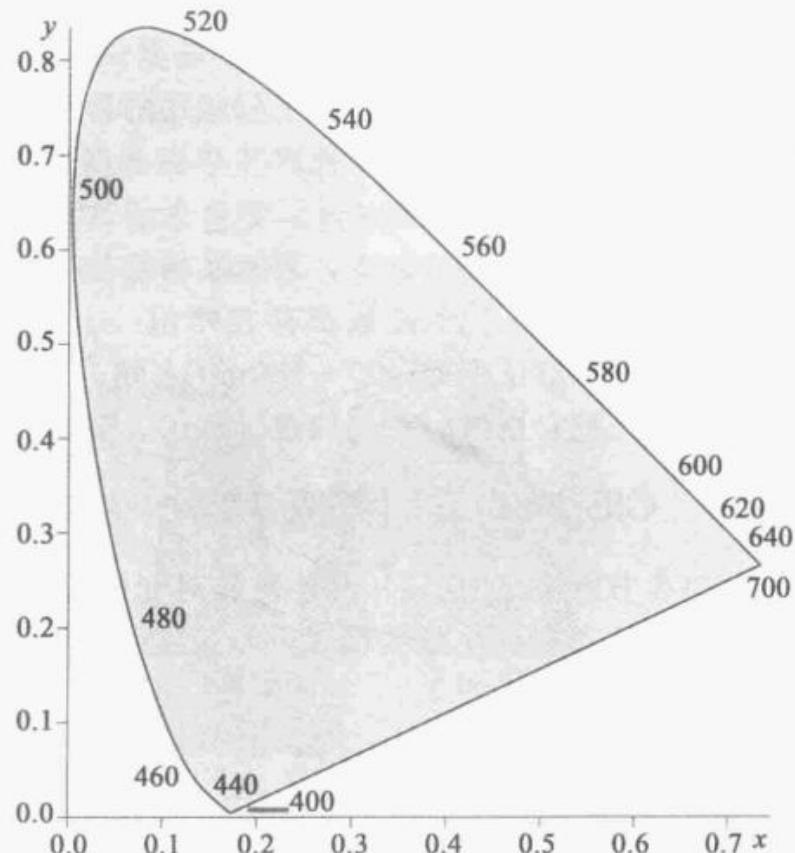


图 25-7 CIE 1931 2° $[x, y]$ 色品图 (色度图)

25.6 CIE $[x, y]$ 色品

出于概念理解上和计算上的方便考虑，定义一种“纯”色，这种颜色没有明度。CIE 还定义了将 XYZ 三刺激值进行归一化，以得到两个色品值 x 和 y 的标准化过程。

色品值是由如下的投影变换计算得到的：

$$x = \frac{X}{X+Y+Z}, y = \frac{Y}{X+Y+Z} \quad (25-1)$$

对第三个色品坐标 z 也做了定义。但是由于 $x+y+z=1$ ，所以 z 值是冗余的。色品坐标 x 和 y 是抽象的数值并且没有直观的物理意义。

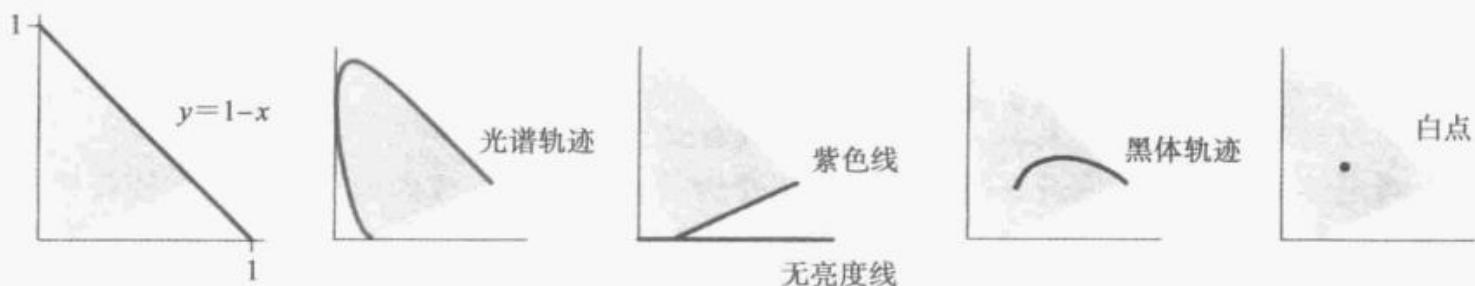
一种颜色也可以由其色品值和照度值， xyY 的形式指定。按式 (25-1) 的逆变换，可以由 $[x, y]$ 色品值和照度恢复出 X 和 Z 的三刺激值：

$$X = \frac{x}{y}Y, Z = \frac{1-x-y}{y}Y \quad (25-2)$$

图 25-7 所示各种颜色在 $[x, y]$ 色品图中都被绘制成点。

如图 25-8 所示，标示出了 $[x, y]$ 色品图的一些特征。其中比较重要的特征处于在直线 $y=1-x$ 上，或处于直线下方、左方。

当某单波长窄带光信号（或者单色光信号）的光谱功率分布集中于 400~700nm 范围内时，它在 $[x, y]$ 平面上大致呈现倒 U 形（或者马蹄形）光谱轨迹。

图 25-8 CIE $[x, y]$ 特征图

紫色不能由单波长的光产生；需要由短波光和长波光相混合才能产生。在色品图中，紫色线从只包含短波光的纯蓝色（紫罗兰）一直延伸到只包含长波光的纯红色。

对于白色光，其实没有明确的物理和感知定义。许多重要照度的光源都是黑体辐射，其色品坐标位于图中的黑体轨迹（有时也称为普朗克轨迹）上。黑体辐射将在下一节中讨论。

如图 25-9 所示，随着温度的升高，绝对功率也随之升高，并且辐射的峰值要向短波方向移动。SPD 呈现为白光时，其 CIE $[X, Y, Z]$ 值约为 $[1, 1, 1]$ ，而且其 $[x, y]$ 色品坐标为 $[1/3, 1/3]$ ；白点位于色品图的中心区域。25.9 节中，将会描绘与白光相关的光谱功率分布。

任何一个恒正的（不论是理论上的还是现实中的）光谱功率分布值都对应于色品图中的一个点，该点处于光谱轨迹和紫色线所围成的区域里。所有的颜色都处于这一区域，而在此区域之外的点则与颜色无关。用是否可见来定义颜色是不合理的，因为颜色本身就是根据视觉定义的，也就是说，如果它不可见，那么它根本就不会称为颜色。

如上文所述，所得到 x 和 y 的投影变换中，两个光谱功率分布或者两个三刺激值的加色法（线性和），在 $[x, y]$ 平面上用直线来表示。然而，距离是不可靠的，所以色品值并不是按照线性的关系混合的。其实不论是 $[X, Y, Z]$ 还是 $[x, y]$ 坐标在实际感知上都不是一致的。

25.7 黑体辐射

马克思·普朗克断定光谱功率分布来源于发热体、黑体辐射物，是被加热物体温度的函数。图 25-9 显示了在几种温度下的黑体光谱功率分布图。随着温度的升高，绝对功率随之提高，同时光谱峰值向短波方向移动。图 25-10 给出的是功率归一化到 555nm 段的曲线。正如图 25-10 所示，如果黑体辐射源的功率在任意一个波长上归一化，那么频谱特征上的动态特征就会很明显。

很多光源的核心都可以看成一个受热

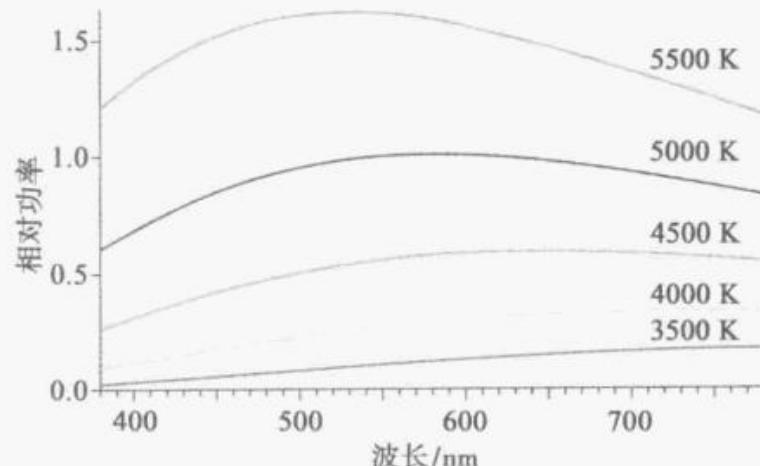


图 25-9 黑体辐射在几个温度点上的 SPD 绘于图中

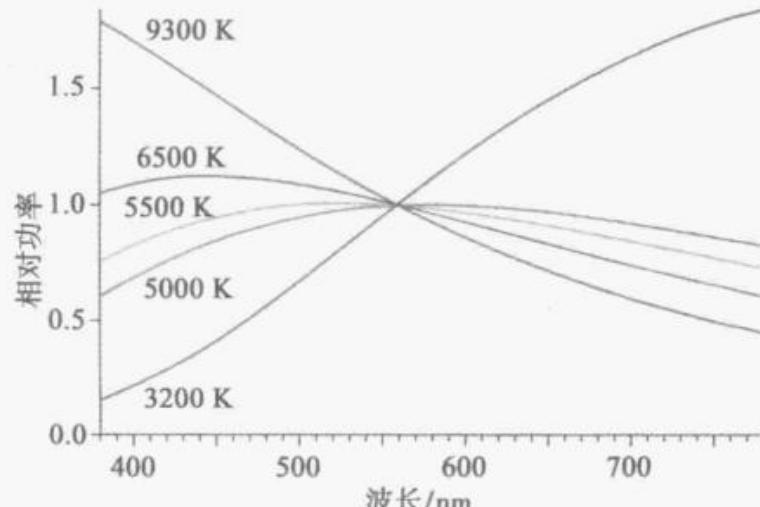


图 25-10 规范化黑体辐射 SPD 图

体，因而，用标定黑体辐射源的热力学温度（单位为开尔文，即 K）来表征其照度水平是很实用的。

开尔文的符号常常写成 K，而不使用度数标记。

当黑体辐射源的温度升高时，黑体轨迹变为 $[x, y]$ 坐标中的跟踪路径。在较低的温度下，该路径呈现出红色（称为红热），当观察者适应 CIE D₆₅ 参考白光时（马上就要讲到）——约 2000K，光源会近似呈现橙色。接近 4000K 时，它呈现黄色。在约 6000K 时，呈现白色。超过 10 000K，就变为蓝色了。

对于色彩学来说，存在一种与之看似矛盾的现象，那就是，冷水龙头通常标为蓝色的，而热水龙头则标为红色。

25.8 色温

光源可以用某种单色光的色温值来表征——与该光源色品相匹配的黑体辐射源的温度。如果这种匹配是近似的，则使用相关色温（Correlated Colour Temperature, CCT）来表示。

CIE 1960 标准 $[u, v]$ 坐标中有所描述。

色温有时也可以用另一个参数扩展，该参数给出了在 CIE 1960 标准 $[u, v]$ 坐标上颜色到黑体辐射轨迹的最近距离——该值就是平常所说的“最小感知色差”（MPCD）。我认为从直观的角度来说，用开尔文结合 $[x, y]$ 和 $[u, v]$ 色品坐标来指定颜色温度更合理。

当黑体辐射源的温度从较低的值（例如 1000K）到较高的值（例如 20 000K）变化时，它在色品坐标系上扫描的路径称为黑体轨迹（见图 25-8）。图 25-8 所示的是以一种不均匀的方式将辐射温度进行了划分。

25.9 白光

如前所述，白色没有唯一的定义：要得到精确的白色，必须指定白光的光谱功率分布和色品特性。正如将会详细叙述的那样，白光点是由等量的红、绿、蓝基色光混合叠加组成的一组三刺激值（或照度和色品坐标）。白光的颜色是基色分量功率的比值函数（在简化的图像重现中，白色由照度和无色媒质的光谱功率分布之积确定）。

有时，为了计算方便起见，将白色定义为其 SPD 功率在整个可见光谱频段均匀分布。这种参考白光称为等能量光源，CIE 标准记为 Illuminant E，它的 CIE 标准坐标为 $[1/3, 1/3]$ 。

还有一种更加现实的参照，就是近似日光，它在 CIE 标准中定义为 D₆₅。在其他方式的优势不是十分明显的时候，应该选用这种参考。在印刷厂中通常使用 D₅₀ 光源，而摄影中使用 D₅₅ 光源。这些表示方法都是室内光（钨丝灯发出的光）和日光的一种折中。图 25-11 表示出了几种标准光源的光谱功率分布情况。对应的色度坐标在表 25-1 中给出。

CIE D 光源用一个两位的数字的下角标来表示。CIE D₆₅ 光源表示的是色温为 6504K 的光源。

很多计算机显示器和消费电视接收器默认色温都设置为 9300K。这种参考光中蓝色的成分比较多，所以更适合于欧洲和美洲国家的电视图像重现。从 D₆₅ 标准到 9300K 的转换过程需要乘以 BT. 909 标准中蓝色刺激值，其数值约为 1.3。然而，对于习惯了亚洲文化的观众来说，默认色温比 D₆₅ 要更偏蓝一些（例如日本的电视演播就是如此）。

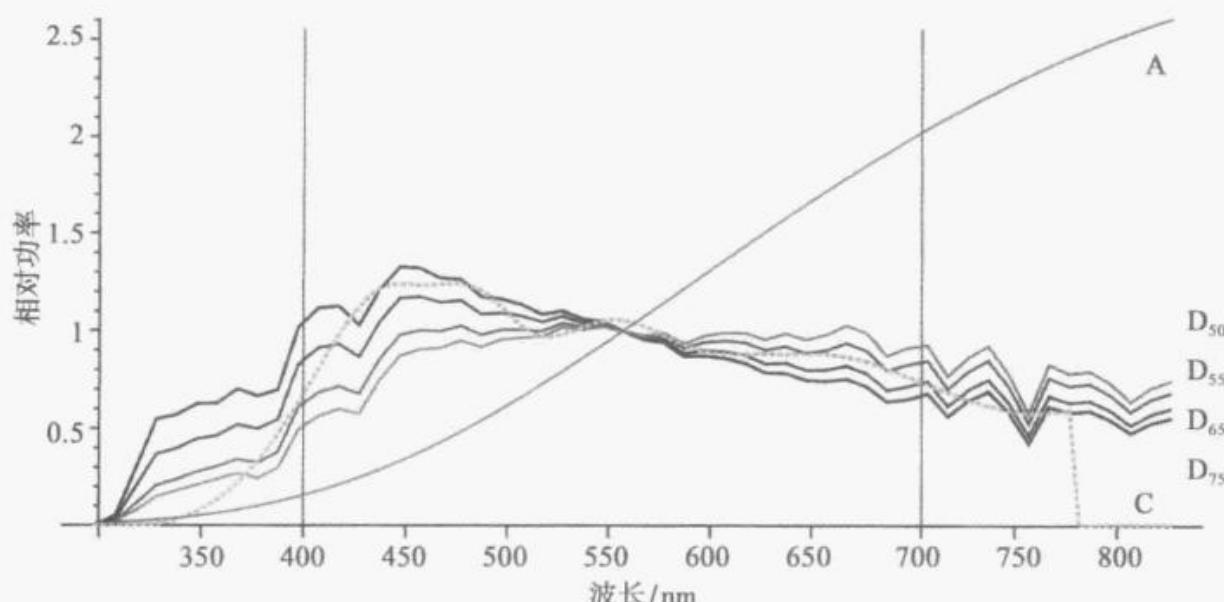


图 25-11 CIE 辐射体

光源 A 是已废弃的钨丝灯光照标准光源。其光谱功率分布近似于 3200K 的黑体辐射曲线，如图 25-10 所示，光源 C 是一个表示日光的早期标准，现在也已经过时了。D 组光源表示几种常见色温的日光光源。

表 25-1 列举出了几种常见参考白光的色品坐标值。其中灰底显示的是较为常用的 D₆₅ 标准的 SD 和 HD 的参考白光。

表 25-1 参考白光

符 号	x	y	z	u'_n	v'_n
1666.7K (6000 Mirek)	0.376 83	0.380 50	0.242 67	0.2213	0.5027
CIE III.A (已废弃), ~2856K	0.447 57	0.407 45	0.144 98	0.2560	0.5243
CIE III.C (已废弃)	0.310 06	0.316 16	0.373 78	0.2009	0.4609
CIE III.D ₅₀	0.3457	0.3587	0.2956	0.2091	0.4882
CIE III.D ₅₅	0.3325	0.3476	0.3199	0.2044	0.4801
CIE III.D ₆₅ , ~6504K	0.312 727	0.329 024	0.358 250	0.1978	0.4683
CIE III.E (相同能量)	0.333 334	0.333 330	0.333 336	0.2105	0.4737
9300K (用于亚洲演播室标准)	0.2830	0.2980	0.4190	0.1844	0.4463
∞ (0Mirek)	0.23704	0.236 741	0.526 219	0.1767	0.3970

有时相对色温的倒数比相对色温更加均匀。摄影师们通常使用单位 Mirek (兆开尔文的倒数 (MK^{-1})), 即 $10^6/t$, 这里的 t 的单位是开尔文 [K]。Mirek 比开尔文这个单位感知上更加均匀。对于标准的视频和摄影获取来说, 相关色温的范围通常为 2000K 到 10 000K, 也就是 500~100Mirek。

Mirek 这个单位有时也叫做兆开尔文的倒数, 曾经称为 Mied (微度数倒数)。

25.10 色彩适应

人类视觉可以适应所看到的环境。我们看到的独立图像, 例如, 一个 35mm 幻灯片或者投射到暗室中的一幅运动图像胶片, 都会产生自己的参考白光。观察者对参考白点的变化十分宽容。然而, 如果同一幅图像使用外界参考白光来观察, 或与另一幅图像一起观看, 那么参考白点的变化将十分明显。完全适应性通常只适用于色温为 5000K 到 6500K 范围。色温 3200K 左右的钨丝灯光源, 则总是显得偏黄一些。

钨丝灯光的色温不能比钨的熔点更高。也就是不能高于 3695K。

25.11 感知均匀颜色空间

正如 3.3 节所述，如果系统中分量值的小波动在该数值范围内的感知结果大致相等，则该系统就视为感知均匀的。

照度并非感知均匀的。在 24.5 节中，描述了照度如何转换为明度，也就是用 L^* 来表示，它是近似感知均匀的，即

$$L^*(Y) = \begin{cases} \left(\frac{116}{12}\right)^3 \frac{Y}{Y_N} & \left(\frac{Y}{Y_N} \leq \left(\frac{24}{116}\right)^3\right) \\ 116\left(\frac{Y}{Y_N}\right)^{\frac{1}{3}} - 16 & \left(\left(\frac{24}{116}\right)^3 < \frac{Y}{Y_N}\right) \end{cases} \quad (25-3)$$

把这个概念延伸到色彩领域，XYZ、RGB 三刺激值和 xyY （色度和照度），都是感知很不均匀的。如何将 XYZ 转换到合理的感知均匀空间上，这一个课题困扰了 CIE 委员会 10 年之久，最终也没有哪个单独系统能够很好解决这个问题。1976 年，CIE 标准化了两套系统，也就是接下来将要讲到的 $L^* u^* v^*$ 和 $L^* a^* b^*$ 。在这两套系统中，感知差值近似为欧氏距离。

$L^* u^* v^*$ 和 $L^* a^* b^*$ 通常写成 CIE LUV 和 CIE LAB。它们通常发音为 SEA-love 和 SEA-lab。颜色参量 $u^* v^*$ 和将要讲到的 $u'v'$ 与 UV 色差参量都没有关系。

25.12 CIE $L^* u^* v^*$ 系统

CIE $L^* u^* v^*$ 的计算首先将三刺激值 $[X, Y, Z]$ 投影成中间量 u' 和 v' 。

$$u' = \frac{4X}{X + 15Y + 3Z}, v' = \frac{9Y}{X + 15Y + 3Z} \quad (25-4)$$

同样， u' 和 v' 也可以通过 x 和 y 色度值算出：

$$u' = \frac{4x}{3 - 2x + 12y}, v' = \frac{9y}{3 - 2x + 12y} \quad (25-5)$$

要从 u' 和 v' 中恢复出 X 和 Z 三刺激值，使用如下的关系式：

$$X = \frac{9u'}{4v'}Y, Z = \frac{12 - 3u' - 20v'}{4v'}Y \quad (25-6)$$

由 u' 和 v' 恢复 x 和 y ，可使用下式：

$$x = \frac{9u'}{6u' - 16v' + 12}, y = \frac{4v'}{6u' - 16v' + 12} \quad (25-7)$$

因为 u' 和 v' 是由投影变换产生的，所以 u' 和 v' 的坐标与 CIE1931 20° $[x, y]$ 色度图是密切相关的。如果想要产生二维的图像可以使用 $[u', v']$ 图像， $[u', v']$ 图能比 $[x, y]$ 图更好地表现色差的感知特性。不过， $[u', v']$ 要乘以 L^* （见式 (25-8)）来生成 $[u^*, v^*]$ 。当照度增加时，这个乘法明显扩大了感知增量的范围。 $[u', v']$ 图的感知差取决于照度水平的变化，但是从图中看并不明显：注意不要由该图得到过于肯定的结论。

CIE 1976 标准规定， $u'v'$ 参量继承 CIE1960 标准的 u 和 v 。两者关系为： $u = u'$ ； $v = 2/3v'$ ，也就是说，1960 标准的 v 参量降低了视觉感知性，通过乘以一个系数值 1.5 来产生 1976 标准的 v' （为了计算得到 1960 标准的 v 值，可以在式 (25-5) 中，将 $9y$ 替换为 $6y$ ）。虽然某种程度上说这是感知均匀的，但是这些元素与 $R'G'B'$ 和 Y' 还是不同的。

要计算 u^* 和 v^* ，应先计算 L^* 。然后由所使用的参考白光 X_N ， Y_N ， Z_N 计算出 u'_N ，

v'_N 。(下标 N 表示规一化)。表 25-1 给出了几种常见参考白光的 u'_N , v'_N 坐标(例如, 无穷大色温的 $[x_N, y_N]$ 坐标为 $[0.237, 0.237]$, $[v_N, v_N]$ 坐标约为 $[0.177, 0.397]$)。最终计算出 u^* 和 v^* 。

$$u^* = 13 \times L^*(Y) \cdot (u' - u'_N), v^* = 13 \times L^*(Y) \cdot (v' - v'_N) \quad (25-8)$$

色域指的是成像系统中的有效色彩范围。对于典型图像再现系统的色域而言, u^* 和 v^* 值的范围均为±100 左右。

$L^* u^*$ 和 v^* 的欧氏距离——表示成 ΔE_{uv}^* , 用来估计色差的感知特性:

$$\Delta E_{uv}^* = \sqrt{(L_2^* - L_1^*)^2 + (u_2^* - u_1^*)^2 + (v_2^* - v_1^*)^2} \quad (25-9)$$

如果 ΔE_{uv}^* 取单位值, 或更小, 色差值就可以认为是不可察觉的。然而, $L^* u^* v^*$ 是感知不均匀的, 只是近似感知均匀。 ΔE_{uv}^* 的值位于 1 到 4 之间时, 不确定能否感知到, 而取决于被检验的色彩空间的区域范围。 ΔE_{uv}^* 值大于 4 就可能感知到, 但也取决于具体情况。

ΔE^* 发音为 delta E-star。

可以用 $[u^*, v^*]$ 的极坐标形式来表示色度和色调:

$$C_{uv}^* = \sqrt{u^{*2} + v^{*2}}, h_{uv} = \arctan \frac{v^*}{u^*} \quad (25-10)$$

此外, 还有一个心理测量饱和的术语:

$$s_{uv} = \frac{C^*}{L^*} \quad (25-11)$$

这里定义的色度、色调、饱和度, 与 HSB、HIS、HSL、HSV、HIS 系统中的定义的用于计算和数字图像处理中的色调和饱和度并不相同: 对这些空间的大多描述和大多公式都与色彩学原理不符。尤其是, 对于明度和三刺激值这两参数来说, 与它在色彩学中的定义是十分矛盾的。

CIE $L^* u^* v^*$ 表现出合理的感知均匀。由于 XYZ、xyZ 和 RGB 是以投影的方式映射到 u' , v' 坐标上的, 所以 $L^* u^* v^*$ 在视频领域得到普及: 这几种空间中的直线仍映射为 $u^* v^*$ 平面中的直线。尽管 $L^* u^* v^*$ 方便实用, 现今的色彩学一致认为 $L^* a^* b^*$ 有更好的感知性能, 将在下节中讲述。

25.13 CIE $L^* a^* b^*$ 参数 (CIE LAB)

参数 $a^* b^*$ 通过下式计算:

$$a^* = \frac{125}{29} \left[L^* \left(\frac{X}{X_N} \right) - L^* \left(\frac{Y}{Y_N} \right) \right], b^* = \frac{50}{29} \left[L^* \left(\frac{Y}{Y_N} \right) - L^* \left(\frac{Z}{Z_N} \right) \right] \quad (25-12)$$

相关系数约为 4.310 和 1.742。这里使用的是非通用的写法: 用 L^* 来代替传统的辅助函数 f 。 L^* 的定义包含具有一阶参数连续性 (C^1 continuity) 非线性段和幂函数段。该非线性区(以 L^* 的形式) 整合成 a^* 和 b^* 参数。

参考量 L^* 从黑到白的范围为 0~100。对于典型标清和高清标准的 BT709 基色来说, a^* 和 b^* 参数的范围分别为 $[-87, \dots, +97]$ 和 $[-108, \dots, +95]$, 并不包含任何高于或低于此范围, 以及非法、无效的 $C_B C_R$ 值。

如在 $L^* u^* v^*$ 标准中所述, $L^* a^* b^*$ 的单位欧氏距离表示成 ΔE_{ab}^* —— 约为色差的感知近似:

$$\Delta E_{ab}^* = \sqrt{(L_2^* - L_1^*)^2 + (a_2^* - a_1^*)^2 + (b_2^* - b_1^*)^2} \quad (25-13)$$

如果 ΔE_{ab}^* 是单位值或更小, 色差就可以忽略不计, 然而, $L^* a^* b^*$ 并不能实现感知

均匀：因为它仅仅是一种近似。

$[a^*, b^*]$ 的极坐标值可以用以下计算公式转换成色度和色调值：

$$C_{ab}^* = \sqrt{a^{*2} + b^{*2}}, h_{ab} = \arctan \frac{b^*}{a^*} \quad (25-14)$$

求得 a^* 、 b^* 坐标的公式不是投影变换： $[x, y]$ 坐标中的直线并不能直接映射成 $[a^*, b^*]$ 坐标中的直线。 $[a^*, b^*]$ 坐标可以画成二维图形，但是这种图形并不是色品图。

25.14 CIE $L^* u^* v^*$ 和 CIE $L^* a^* b^*$ 总结

这二者都将 XYZ 的感知不灵敏性由 80:1 提高到 6:1 左右。这两个系统都将三刺激值变换为 0~100 的明度分量，且两个亮度分量的范围均为 ±100 左右。两个系统的单位欧氏距离都近似为颜色的最小可觉差。

考虑到 L^* 的范围是 0~100， u^* 和 v^* 的范围均为 ±100 左右。 ΔE_{uv}^* 在单位值以内可以定义四百万种颜色，其中一百万种颜色可以被视觉所分辨，所以 CIE $L^* u^* v^*$ 提到的颜色数还是相对保守的。一百万种颜色——甚至于用 ΔE_{uv}^* 和 ΔE_{ab}^* 可分辨的四百万种颜色——在 24 位真彩色系统中很好呈现，该系统使用感知近似变换函数（例如 BT.709 的函数）可呈现 1670 万色（不过，每个像素点 24 位对于充分体现线性光编码的性能来说还是显得太少了）。

McCamy 认为正常情况下，大约 187 500 种颜色可以分辨出来。请参考 McCAMY, CAM S. (1998) 23 章 (5) 中的“可分辨颜色”部分。

$L^* u^* v^*$ 和 $L^* a^* b^*$ 系统主要应用在色彩规范中。虽然这二者都成功应用于静态图像编码——尤其是打印系统中，但是对于经济型实时视频处理来说，这两种系统需要的计算复杂度都太高了。它们的计算复杂度使其不适用于图像编码。视频编码使用的 $R'G'B'$ 非线性编码具有很好的感知均匀性能，也很利于实时处理。需要注意的是， $R'G'B'$ 通常包含特定的色域限制，而 $L^* u^* v^*$ 和 $L^* a^* b^*$ 系统则能表示所有的颜色。 $L^* a^* b^*$ 系统有时也用于计算机图形处理， $[a^*, b^*]$ 坐标的范围为 -128~+127（例如在 Photoshop 软件中）。虽然有这些限制，CIE $L^* a^* b^*$ 系统还是几乎涵盖了所有颜色。

25.15 色彩规范和彩色图像编码

色彩规范系统需要能以很高的精度表示任何的颜色。因为每次只能处理很少的颜色，所以这样的系统计算是很复杂的。色彩规范系统必须与 CIE 系统密切相关。色彩规范中比较有用的系统有：CIE XYZ 以及由它延伸出来的 xyY 、 uv 、 $L^* u^* v^*$ ，以及 $L^* a^* b^*$ 系统。

一个彩色图像可以用一个像素矩阵来表示，每个像素都包含定义颜色的三个值。正如读者在本章中所了解的，三个分量足够表示任何颜色了（在打印系统中，添加第四个分量很方便，即黑色分量，表示为 CMKY）。

理论上来说，图像编码的三刺激值可由色彩规范系统提供。然而，实际的图像编码系统必须足够高效，并不能做到无限精确，也不需要跟 CIE 系统联系得那么紧密，通常只需要覆盖一个合理的颜色范围即可，而不需覆盖所有可能的颜色。所以，颜色编码会用到多种与颜色规范不同的系统。

利于图像编码的系统是线性 RGB 系统。非线性 RGB 系统（通常表示成 $R'G'B'$ ，或 sRGB 看成一个变量），非线性 CMY 系统、非线性 CMYK 系统，以及 $R'G'B'$ 派生出的系

统, 例如, $Y'C_B C_R$ 、 $Y'P_B P_R$ 等这些系统都总结在图 25-12 中。图中色彩系统可以分为四组, 各组之间有不同的变换关系。三刺激值系统、感知均匀系统, 都利于图像编码 (图中的 HSB、HIS、HSL、HSV 和 HIS 符号用问号进行了标记, 是因为这些系统缺少颜色的客观定义)。

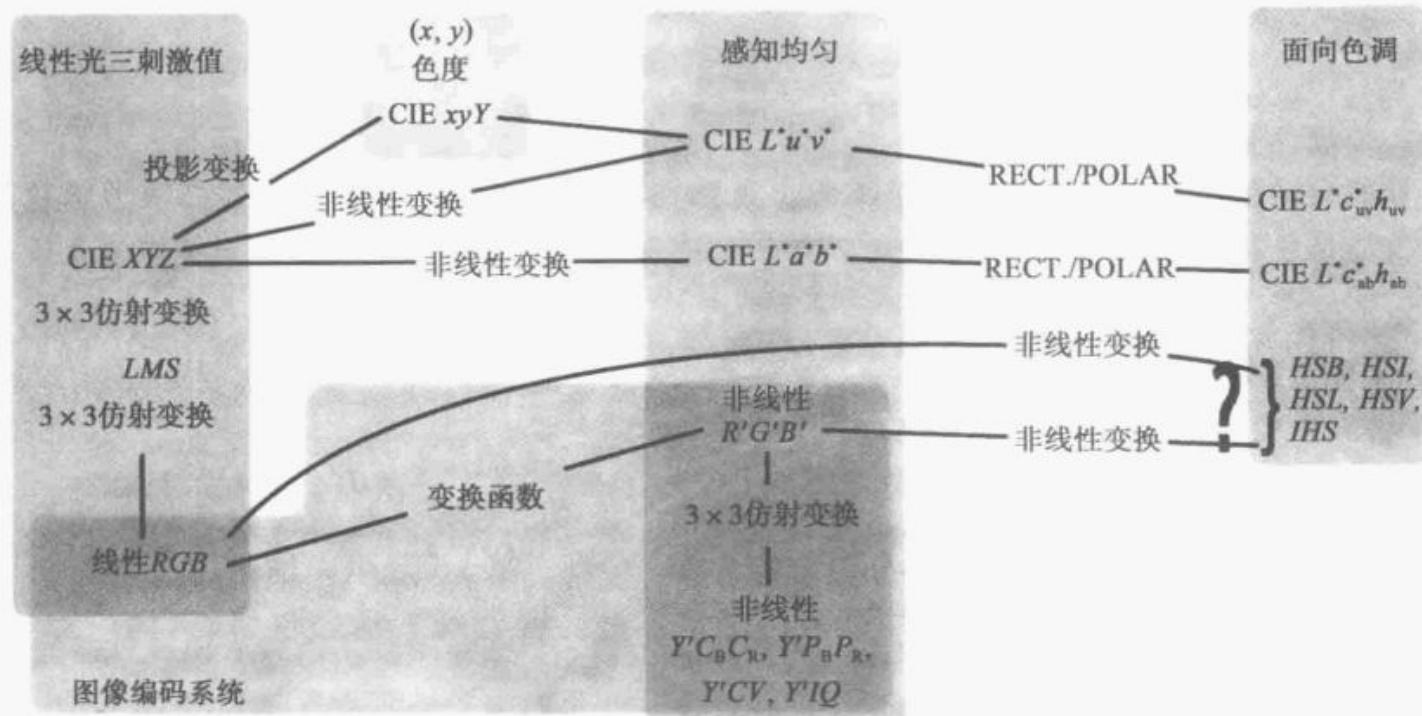


图 25-12 色彩系统

如果在汽车制造领域, 那么必须将车门上的油漆与挡板上的油漆相匹配, 这就需要颜色规范了。这里计算量不是太大, 因为只有两个颜色元素, 车门和挡板。要传送一副汽车的图片, 可能有一百万个或更多的颜色元素, 因而计算必须非常高效, 这就需要图像编码系统。

延伸阅读

Wyszeck 和 Stiles 所著的“色彩科学”是色彩学中的圣经, 但是它着实令人望而生畏。因为它所涉及的颜色都是通用的, 几乎不包含专用于成像的相关信息。

若想看到更通俗易懂的色彩理论介绍, 以及图像重现的实用描述, 可以参考 Hunt 所著的经典著作。

由 Bern 修订的 Billmeyer 和 Saltzman 的经典著作, 对色彩学做了很好的介绍。如果要用非数学计算的角度去研究色彩学和感知理论, 可以参阅 Rossotti 著的书。

第 26 章

视频色彩学

传统色彩学，已经在前面章节中讲述过，它奠定了现代数字色彩学的基石。然而，色彩学的目的是规范颜色而不是图像编码。虽然在视频中获得好的颜色性能必须理解色彩学，但是色彩学的严格应用却是不切实际的。本章将介绍在制作实际照相机和编码系统时要做的一些工程应用的折中。

视频处理一般关心的是用来源于具体场景的三种分量来表示颜色，通常用红、绿、蓝或者由此衍生出来的分量表示。能否精确重现颜色取决于原始场景中物理光谱如何变换为分量值，以及分量值如何反变换为显示的物理光谱。这些都是本章的主要内容。

一旦获得了某个场景的红色、绿色和蓝色分量值，这些值将转换为其他更适合于处理、录制和传输的形式。这部分内容将在第 29 章、第 30 章中详细讲述（虽然目前 BT. 709 标准基色在高清和标清中是通用的，不过，两者的其他颜色编码内容是有分别的）。

在之前的章节中，已经讲述了怎样将场景元素的 SPD 转换为表示颜色的 XYZ 三刺激值。一种直接呈现这些颜色的方式是让显示系统再现这些 XYZ 值。这种原理在诸多色彩重现过程中都有所应用，这也是视频色彩理论的基础。然而，在图像再现中，鉴于感知方面的原因，直接再现 XYZ 值并不合适。必须进行一些修正才能得到可接受的结果，这些修正有关的内容在第 10 章做了讲解。

读者可能在考虑是否该跳过本章，需要注意的是，用 RGB 坐标表示颜色的精确度主要取决于 RGB 特征基色和它们的功率比（与参考白光的比）。如果系统自始至终使用的是固定的基色集合，就像在标清和高清系统中一样，那么就不需要考虑 RGB 系统的不同特性。

然而，如果图像在不同的阶段或者产品有不同的基色集合，比如，在数码成像或者数字静态摄影表示中——那上述特性就至关重要。

26.1 加性重建 (RGB)

在之前的各章节中，曾解释了物理的光谱功率分布如何分析成表示颜色的三种基色。这一章的内容是解释这些基色如何组合在一起表示（或者说重现）颜色。

重现一定范围色彩的最简单方法是像图 26-1 所示的那样混合三种不同颜色的光束。从物理学来说，每一种光的频谱相加到一起形成了混合光的频谱。从物理和数学的角度来说，这属于谱叠加，整个过程称为加性重现。

这里所说的 RGB+W 显示方式是在 20 世纪 90 年代和 21 世纪初提出来的。其最初主要是应用在连续颜色数字光处理 (DLP) 投影机上的。在这种显示方式中，假定白光的照度比红绿蓝三种颜色的总照度之和还要多，要求有高的明亮度。然而，这种显示不能扩展到额外的颜色混合中去，正如本文所写到的，其实几乎所有的展示设备，包括图形图像，消费需求颜色描述，RGB+W 的投影机等都不再受到偏爱了。

在本书，描述了 Grassmann 第三定律：颜色视觉服从叠加定律，即由任何三种基本光

谱功率分布产生的加性合成颜色都可以通过将 XYZ 基色的三刺激值分量相应的数值相加来预测。可由特定 RGB 基色集合形成颜色完全由各个基色的颜色（三刺激值、或照度值和色品坐标）确定。在图像学、照相学、商业打印中使用减性重建，就要复杂得多了：因为减性重建中的颜色并不由各个基色的颜色确定，而是由其光谱特性决定。

如果读者对这里的照度这个词或者标志 Y、Y' 不熟悉，可以参照第 24 章。

图 26-1 给出了视频中使用的彩色加性混合物理过程示例。每个基色都有对图像独立、直接的路径。图像的功率谱是这些基色光谱的和。合成图像完全由组成它的基色决定，所以对合成图像的分析和预测都很简单。这里举的例子是索尼电子管显示器的光谱功率分布（SPD）。

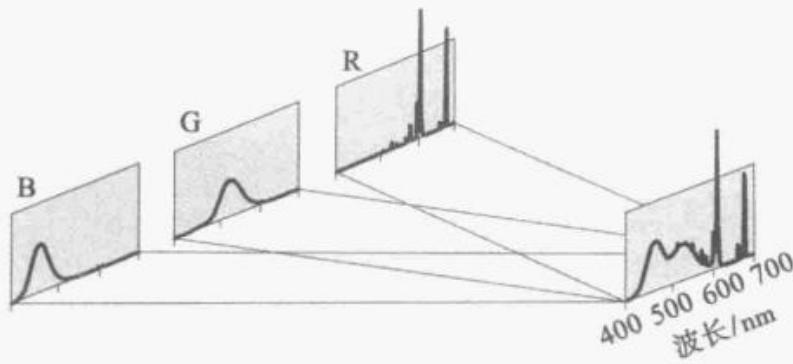


图 26-1 加性重建

在加性重建直接应用于视频投影仪时，红绿蓝光束的光谱在投影机屏幕表面直接进行物理相加。加性重建也应用于直接显示颜色的 CRT 显示器中，仅有一些间接的处理。CRT 显示器屏幕上有一些小的荧光点，当它们被对应的光束打击时，就产生了红、绿和蓝光。这时，从一定距离去观察屏幕时，这些小点就投射到了观察者的视网膜上。单独呈现红绿蓝光的基色将产生最宽范围的颜色。当颜色显示系统只是 CRT 时，RGB 特征被它们的荧光粉的色度所限制，称为荧光粉色品特性。对于不使用荧光粉的新设备，使用“基色品”这个词来代替。

合理选择的三种基色可以产生大范围的颜色，但是有限的基色并不能产生所有的颜色。因而必须进行经济上的折中，在覆盖颜色范围的广度和基色尽可能少（一般选择三个基色）之间做出权衡和取舍。

26.2 RGB 基色特性

加性 RGB 系统的特性由其基色和其白光参考点的色度所决定。如果并不清楚 RGB 图像的基色色度情况，那么就无法精确重现该图像。图 26-2 画出了一些将要讨论的 RGB 系统的基色。

BT. 709 系统指定了高清系统的基色。BT. 709 系统基色三角如图 26-2 的阴影部分所示。

某一范围的颜色或者说全色域，可由给定一组 RGB 基色得到，在 $[x, y]$ 色品图中用一个端点是颜色基色的三角形来表示。虽然 $[x, y]$ 的取值范围可以包含在一个照度函数中，但是这个二维图形并不能揭示出全部原理。例如，位于 $[0.15, 0.06]$ 的 BT. 709 标准蓝色仅仅在照度水平低于白光的 7% 时可以达到。在参考白光下，没有任何色度的偏移。全色域必须考虑所有的三个维度，全色域将在本书继续讨论。

从计算的角度来说，sRGB 标准目前使用十分广泛，它使用 BT. 709 标准中的基色。在桌面计算机的许多应用中，如果图像没有其他附加信息的话，那么默认为使用 sRGB 标准。

图 26-2 重叠画出了 SMPTE/DCI P3 标准基色到 D-cinema 的标准化过程。BT. 709 是高清视频的国际标准，可以合理地用于标清视频中。它包含了 CIE D₆₅ 白光参考点。SMPTE/DCI P3 标准用于 D-cinema 中，其白光参考点位置为 $[0.314, 0.351]$ 。

以下将对这些系统逐个详细介绍。

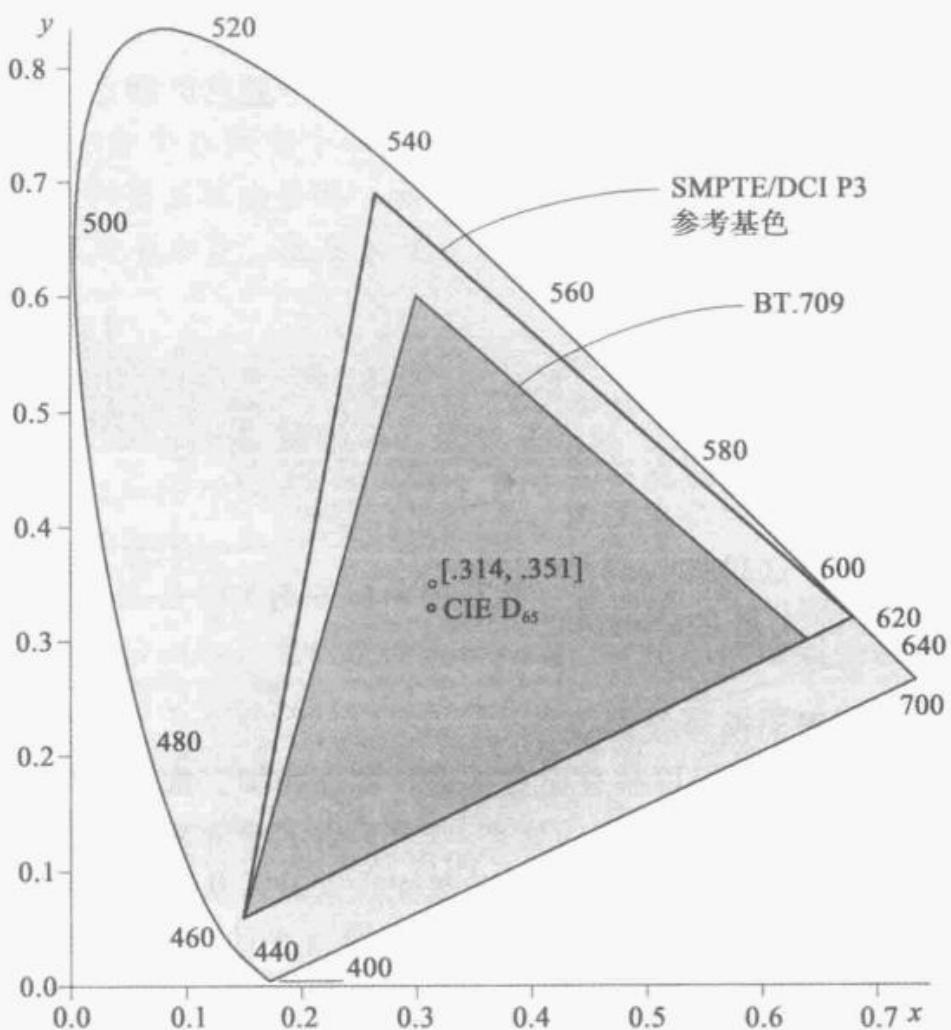


图 26-2 BT. 709 和 SMPTE/DCI P3 基色比较

26.3 BT. 709 基色

ITU-R（前身为 CCIR）组织在 1990 年达成了关于高清晰度电视系统中所使用基元的相关共识。该标准正式表述为 ITU-R BT. 709 建议标准（前身为 CCIR Rec. 709 标准），本书简称其为 BT. 709 标准。这虽然听起来有些不合情理，但其实 BT. 709 标准是某种政治妥协的结果，这个过程中涉及 EBU 红色、EBU 蓝色，以及 EBU 绿色和 SMPTE 绿色的平均构成的绿色。这些基色应用于 sRGB 标准中，便于计算和计算机绘图。BT. 709 标准的基色在现在演播室视频中很具代表性。表 26-1 表示了 BT. 709 标准的基色和其 D₆₅ 白光参考点的色品特性。

ITU-R 的 Rec. 709 标准，规定了视频和国际交流所用的高清数字电视标准。

表 26-1 所示的 BT. 709 基色，应用于 1280×720 像素和 1920×1080 像素高清视频系统，以及个人计算机的 sRGB 标准。

表 26-1

	红 色	绿 色	蓝 色	白色, D ₆₅
x	0.64	0.3	0.15	0.312 727
y	0.33	0.6	0.06	0.329 024
z	0.03	0.1	0.79	0.358 249

表 26-2 给出了相对照度 (Y) 和 BT. 709 色彩空间中色条的 [x, y] 色品值。

表 26-2 BT. 709 色条的照度和色度

	白 色	黄 色	青 色	绿 色	品 红	红 色	蓝 色	黑 色
Y	1	0.927 825	0.787 327	0.715 152	0.284 848	0.212 673	0.072 175	0
x	0.312 727	0.419 320	0.224 656	0.3	0.320 938	0.64	0.15	不定
y	0.329 023	0.505 246	0.328 760	0.6	0.154 190	0.33	0.06	不定

当除数 $X+Y+Z$ 趋近于 0 时, x 、 y 值会突然激增, 如表 26-2 所示, 黑色对应列中 x 和 y 的值为不定。黑色将明显覆盖整个的 $[x, y]$ 图。

视频标准指定了与实际显示器紧密匹配的 RGB 色度。物理显示设备具有误差和不确定性, 但如果显示设备在一定容错范围内遵循 BT. 709 标准, 则可以认为显示器是与设备无关的。

作为演播室视频、广播电视、高清视频的交换标准和固定感知基础的一种标准, BT. 709 确保了其参数在与 CRT 显示器不同的平面显示器中仍可继续使用。然而, 毫无疑问的是, 不断涌现的新型显示器肯定会提供范围更宽的色域。美国电影电视工程师学会 (SMPTE) 采用了一种适用于数字摄像的标准, 后面将会讲到。SMPTE/DCI P3 标准提供了比 BT. 709 标准更加广阔的色域。然而, 使用 P3 本地色彩空间的数字电影不太可能提供给消费者。IEC 61966-2-4 ($xvYCC$) 初衷是给消费视频提供更加宽广的色域, 但是由于缺乏特定的色域映射算法, 所以让人十分怀疑这些初衷最终是否会实现。

26.4 老式标清基色

1953 年, NTSC 确定了紧急彩色电视基色, FCC 对这些基色进行了标准化, 不过现在都已经弃用了。历史上 RP145 (SMPTE-C 标准) 基色曾应用在北美和日本的标清系统中。EBU 基色则用在欧洲的标清系统中。

这些颜色集合的色品特性如表 26-3~表 26-5 所示。

表 26-3 NTSC 基色^①

	红 色	绿 色	蓝 色	白色, CIE III.C
x	0.67	0.21	0.14	0.310
y	0.33	0.71	0.08	0.316
z	0	0.08	0.78	0.374

^①已废弃, 曾于 1953 年起用于 480i 的标清系统, 直到约 1970 年, SMPTE RP145 标准提出的基色被采用为止。

表 26-4 EBU Tech. 3213 基色用于 576i 标清系统

	红 色	绿 色	蓝 色	白色, D ₆₅
x	0.640	0.290	0.150	0.3127
y	0.330	0.600	0.060	0.3290
z	0.030	0.110	0.790	0.3583

表 26-5 SMPTE RP 145 基色用于 480i 标清系统

	红 色	绿 色	蓝 色	白色, D ₆₅
x	0.630	0.310	0.155	0.3127
y	0.340	0.595	0.070	0.3290
z	0.030	0.095	0.775	0.3583

将 RP145 RGB 值理解为 BT. 709 的 RGB 值会引起多达 $20 \Delta E$ 的颜色误差，平均误差约为 $14 \Delta E$ 。EBU 标准和 BT. 709 标准的基色色度不同，将 EBU RGB 值理解为 BT. 709 的 RGB 值将引入最多 $5 \Delta E$ 的颜色基色偏移，其平均偏差约为 $2 \Delta E$ 。

虽然实际标清和 BT. 709 标准存在这些误差，但是 BT. 709 标准基色通常能有效地转换为标清系统，只是转换速度慢一些。

26.5 sRGB 系统

sRGB 标准在计算机中应用广泛。BT. 709 标准基色已经包含在 sRGB 标准中。需要注意的是，虽然基元色品是相同的，sRGB 标准传递函数却与视频系统中标准化的传递函数略有不同。

在 20 世纪 80 年代到 90 年代，计算机中的 RGB 图像数据经常不携带基色色度、参考白光或者传递函数（将第 27 章中讲述）等的有效信息来互相转换。如果 RGB 图像不包含这些参数，那就不可能精确重现这些图像。进入 21 世纪以来，sRGB 标准得到了快速发展，如今，如果 RGB 图像没有上述标识，将其根据 sRGB 标准来解释数据是很安全的。

26.6 SMPTE 无标度基色

2011 年，SMPTE 制定了一个体系，名为无标度色域（Free Scale Gamut, FS-Gamut），该体系包含 2k 和 4k 数字影院产品覆盖的宽色域图像数据（参见 26.19 节）。体系定义了一组默认基色，这些基色并非物理性基色，因而可以超过原始相机负片的正常色域。其中白光参考点符合 D₆₅ 标准。该体系初衷是应用于参考场景的图像编码，在相同的标准中，它通常使用类似的指定编码。默认的 FS 基色和参考白光在表 26-6 中给出。

SMPTE ST 2048-1, 2048×1080 像素和 4096×2160 像素数字电影制作的图像格式 FS/709。符号 FS/709 意欲表明该标准适用于 FS-Gamut 编码和 BT. 709 编码的图像。

表 26-6 SMPTE “Free Scale” 默认基色^①

	红 色	绿 色	蓝 色	白 色
<i>x</i>	0.7347	0.14	0.1	0.312 72
<i>y</i>	0.2653	0.86	-0.029 85	0.329 03
<i>z</i>	0	0	0.929 85	0.340 65

^①超出了传统获取、处理和投影的运动图像胶片的色域范围，图像数据是参照场景生成的。

26.7 AMPAS ACES 基色

运动图像艺术科学委员会（AMPAS）科学技术分会，就数字影像获取和处理中使用的基色达成相关一致。ACES 标准的绿色和蓝色是非物理性的，这样做目的是可以超出通常处理和打印的相机负片的色域范围，其参考白光为 D₆₀。该系统意在对现场参考的图像进行编码。ACES 基色的色度和白光特性在表 26-7 中给出。表 26-7 AMPAS ACES 基色超出了通常获取、处理和投影运动图像胶片的色域范围。ACES 数据是参照场景生成的。

有些读者可能会对 ACES 标准中蓝色坐标 *y* 的负值感到疑惑：当蓝色值增加时，实际上照度是降低的。

表 26-7 AMPAS ACES 基色

	红 色	绿 色	蓝 色	白 色
x	0.7347	0	0.0001	0.32168
y	0.2653	1	-0.0770	0.33767
z	0	0	1.0769	0.34065

26.8 SMPTE/DCI P3 基色

2000 年，数字电影倡导组织（DCI）的参与者们就参考投影使用的基色达成了一致，这些参考投影基色粗略地指定了运动图像的色域范围。图像数据是显示参考的。DCI 协议已经作为 SMPTE 标准发布。该标准的参考白光与现行的实际电影标准是匹配的。在参考白光处的误差在 D61 标准中有所说明。DCI 基色的色域和参考白光在表 26-8 中给出。表 26-8 SMPTE/DCI P3 基色粗略地表示了数字图像使用的色域范围，P3 数据是显示参考。

表 26-8 SMPTE/DCI P3 基色

	红 色	绿 色	蓝 色	白 色
x	0.680	0.265	0.150	0.314
y	0.320	0.690	0.060	0.351
z	0	0.050	0.790	0.340

在针对数字电影的图像数据编码标准建立的过程中，DCI 寻找一种不会过时的标准，该标准能够表示更宽的色域范围，并且不需要表示数据的负值和高于单位值的数据，同时也不需要保留元数据。该标准满足了图像数据编码成显示 XYZ 基色的需求。XYZ 编码使用色度是 $[1, 0]$, $[0, 1]$, $[0, 0]$ 的基色，其参考白光的色度为 $[1/3, 1/3]$ 。

尽管 DCI 采用 XYZ 编码，所有当今使用的数字电影设备都使用 DCI P3 色域范围，如前所述。我认为，这种现状还会持续 5~10 年。如果在传统显示设备中使用更新、更宽色域的材料，那么色域映射将最终得到更新。DCI 标准并没有提到色域映射这部分内容。

26.9 CMF 和 SPD

读者可能在想，是否可以实现某种显示器的基色功率谱分布与 CIE 系统的谱分布曲线——也即与 XYZ 配色函数有着相同的形状。这种显示器可以实现的，但是由 XYZ 三刺激值驱动时，并不能正确重现相应的颜色。而由 XYZ 三刺激值驱动时，也有可以精确再现颜色的显示基色，但是这些基色的功率谱形状与 $\bar{x}(\bar{\lambda})$, $\bar{y}(\bar{\lambda})$, $\bar{z}(\bar{\lambda})$ 的配色函数（CMF）的形状有所不同。要明白其中的原因，就需要对色彩捕捉和再现有明确和清晰的理解。

要找到根据 XYZ 三刺激值再现颜色所需要的一组基色，就需要建立三个功率谱分布函数，也即由 $\bar{x}(\bar{\lambda})$, $\bar{y}(\bar{\lambda})$, $\bar{z}(\bar{\lambda})$ 配色函数进行分析，并分别生成 $[1, 0, 0]$, $[0, 1, 0]$, $[0, 0, 1]$ 。 $\bar{x}(\bar{\lambda})$, $\bar{y}(\bar{\lambda})$, $\bar{z}(\bar{\lambda})$ 的配色函数在整个光谱段上是恒正的。要生成 $[0, 1, 0]$ ，就要求所需基色的光谱功率分布（SPD）内的某些波长为正分布，然后，才可以处理。然而，并没有波长分量对于 Y 值有贡献，而对 X、Z 值也没有明显的正向贡献。

面对这一困境，一种解决的方法是将 X 和 Z 强制置为 0，这可以通过将某些波长区对应的功率谱分布值设为负值来实现。虽然这在数学和信号处理上是没有问题的，但是对于光传感设备来说，负值功率谱是物理不可实现的，因为光功率不能输出一个负值。所以并

不能构建这样一个与 XYZ 直接对应的实际的显示系统。但也正如读者将会看到的那样，对于负光谱功率分布的概念、非物理光谱功率分布和不可实现基色在理论和实际中是非常有用的。

为了理解颜色变换的数学细节，正如本节中所描述的那样，读者应该对线性代数（矩阵）知识比较熟悉。如果读者对线性代数并不熟悉，可以参考 STRANG、GILBERT（1998 年）所著的《Introduction to Linear Algebra》（第 2 版）。

有很多的方式可以选择与 $\bar{x}(\lambda)$ 、 $\bar{y}(\lambda)$ 、 $\bar{z}(\lambda)$ 配色函数相对应的非物理可实现的基色光谱功率分布。其中一种方式是任意选择三种显示基色，其功率分别集中在三个不同的离散波长上。考虑三种显示光谱功率分布，其中每一个都在 600、550、470nm 处有一定的功率值。对前面章节中给出的矩阵中的 $\bar{x}(\lambda)$ 、 $\bar{y}(\lambda)$ 、 $\bar{z}(\lambda)$ 函数分别在这三个波长上进行采样，参见 25.5 节。由此得到的三刺激值如表 26-9 所示，表 26-9 用于解释准确色彩重现中信号处理必要性的基色示例。

表 26-9 示例基色

	红, 600nm	绿, 550nm	蓝, 470nm
X	1.0622	0.4334	0.1954
Y	0.6310	0.9950	0.0910
Z	0.0008	0.0087	1.2876

这些系数可以用一个矩阵来描述，该矩阵中列向量给出了与显示器中的纯红、纯绿、纯蓝对应的 XYZ 三刺激值，即 $[1, 0, 0]$ 、 $[0, 1, 0]$ 、 $[0, 0, 1]$ 。在该矩阵中用一个缩放因子很容易使中间行的和为单位值，因为我们只是想实现相对匹配，而不是绝对匹配。那么有

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.618\ 637 & 0.252\ 417 & 0.113\ 803 \\ 0.367\ 501 & 0.579\ 499 & 0.052\ 999 \\ 0.000\ 466 & 0.005\ 067 & 0.749\ 913 \end{bmatrix} \cdot \begin{bmatrix} R_{600\text{nm}} \\ G_{550\text{nm}} \\ B_{470\text{nm}} \end{bmatrix} \quad (26-1)$$

该矩阵是基于非典型的功率谱分布 RGB 分量，对于典型 RGB 值，请参见式 (26-8)。该矩阵给出了从 RGB 值到 XYZ 值的转换。我们更关心其逆过程，也就是从 XYZ 到 RGB 值的转换，所以，这里给出其逆变换为：

$$\begin{bmatrix} R_{600\text{nm}} \\ G_{550\text{nm}} \\ B_{470\text{nm}} \end{bmatrix} = \begin{bmatrix} 2.179\ 151 & -0.946\ 884 & -0.263\ 777 \\ -1.382\ 685 & 2.327\ 499 & 0.045\ 336 \\ 0.007\ 989 & -0.015\ 138 & 1.333\ 346 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (26-2)$$

式 (26-2) 的矩阵列向量给出了，对于每个基色，需要显示单位 XYZ 三刺激值对应这三个离散波长的权值。CIE XYZ 的配色函数如图 26-3 所示。图 26-4 所示的是与这些函数对应的基色光谱功率分布集。正如预期的那样，显示基色有很多负值光谱分量：基色的光谱功率分布是物理不可实现的。由 XYZ 三刺激值再现图像的任意基色组合都必须是过饱和的，且比任何可实现的光谱功率分布饱和度更高。

要确定一组由 XYZ 再现图像所需的物理可实现的光谱功率分布，可以从另一个角度来思考这个问题：给出一组物理可实现的显示基色，什么样的 CMF 适合直接使用这些基色的混合来再现图像呢？这里关联 RGB 分量和 CIE XYZ 三刺激值的矩阵元素值是恒正的。但是分析这些场景需要的 CMF 存在负值：这种分析滤波器是不可实现的。

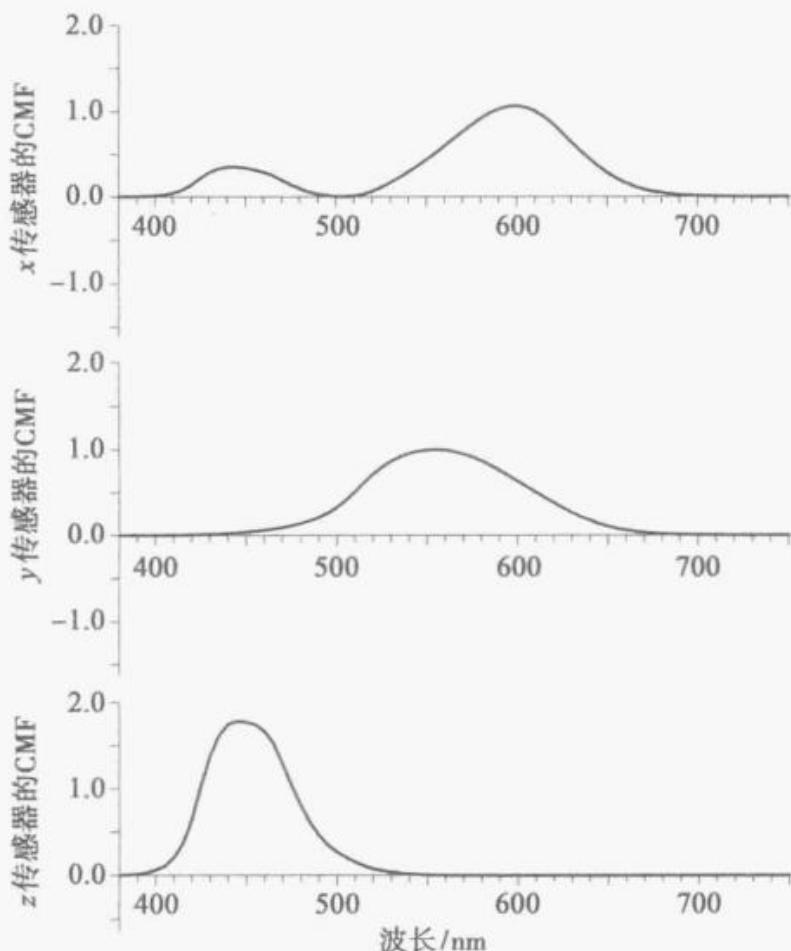


图 26-3 CIE XYZ 三基色的 CMF

要获取某个场景中的所有颜色，则需要有 $\bar{x}(\lambda)$ 、 $\bar{y}(\lambda)$ 、 $\bar{z}(\lambda)$ 光谱灵敏度的滤光器。这些函数是非负的，因此，实际中可以实现。然而，基于很多工程上的考虑，这些函数在实际的照相机或扫描仪中很少使用。

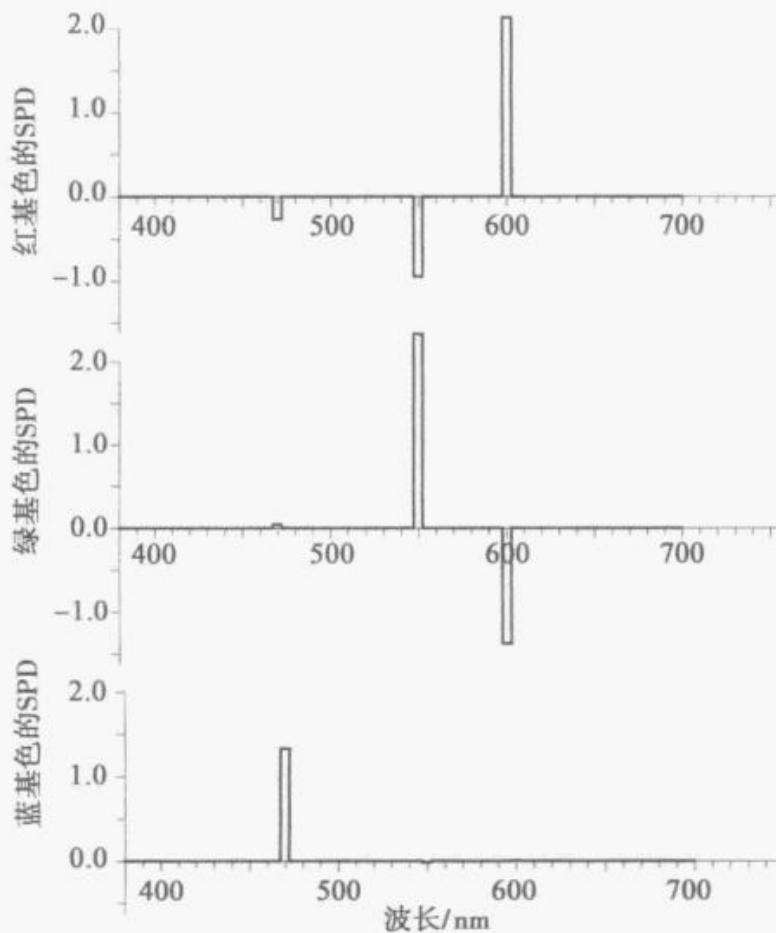


图 26-4 CIE XYZ 三基色的 SPD

要直接再现已经用 CIE XYZ 配色函数分析过的场景，需要非物理性并且含有负值的基色，但这些基色在实际中是不可实现的。有多种不同的可能的基色组合。在这一假设的例子中，在每个基色处的功率集中在相同的三个离散波长处，即 470、550、600nm 处。

图 26-5 给出了对应的配色函数。正如预期，配色函数存在负项，因而不能直接实现。不过，这都是理想化的配色函数（CMF），或者理想的 BT. 709 基色的特点。

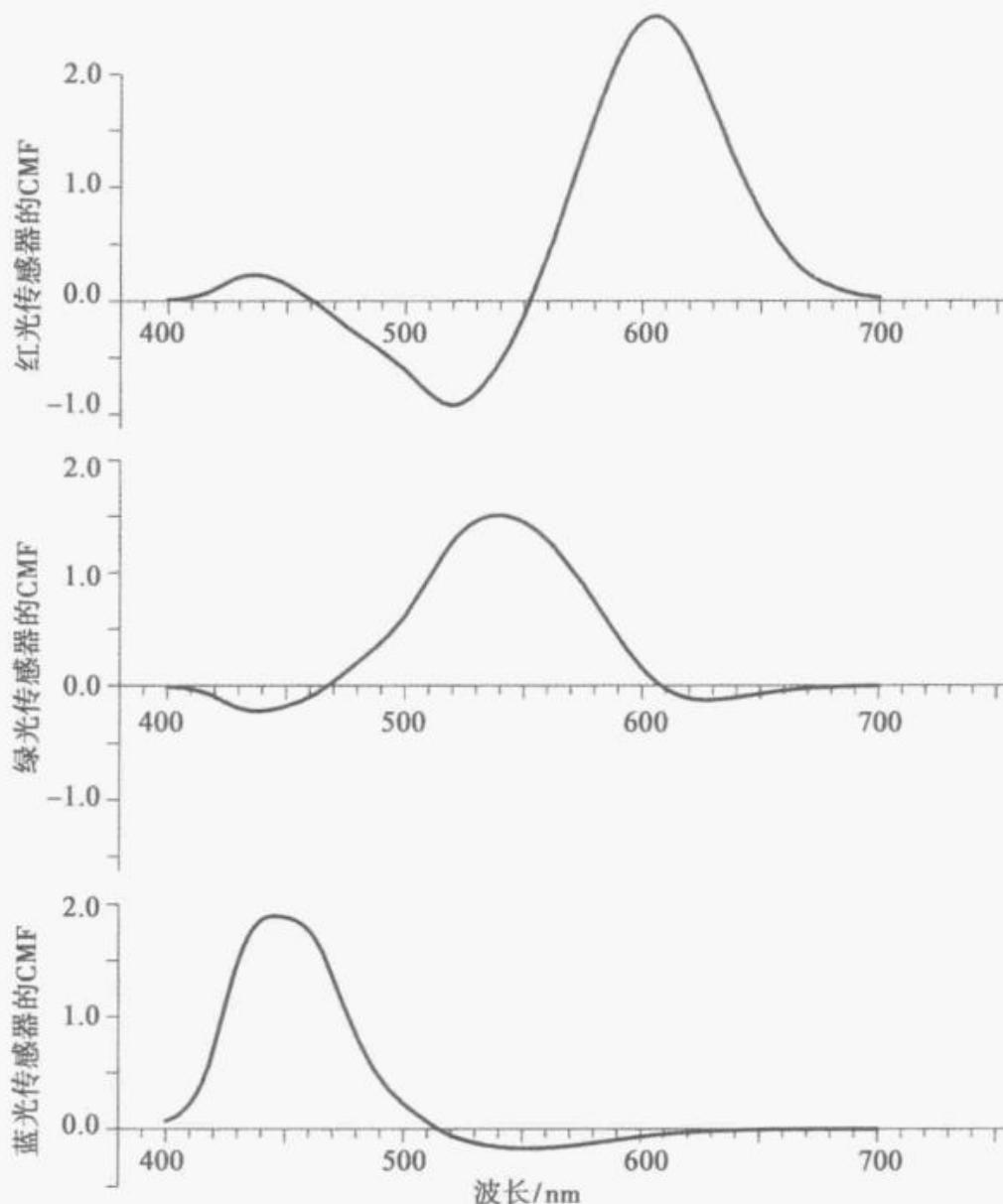


图 26-5 BT. 709 三基色的 CMF

这些分析函数理论上可正确获取用 BT. 709 基色显示的 RGB 分量值。但由于这些函数存在负值，所以不能在相机或扫描仪中直接实现。不过，它们可以通过对图 26-3 所示的配色函数进行 3×3 矩阵变换来实现。

图 26-6 给出了一组转换为 SMPTE 240M 的基色光谱功率分布，与 BT. 709 标准的类似。很多不同的光谱功率分布能够与这些色度产生精确的匹配。图 26-6 给出的基色组是基于索尼单枪三束显示器上的。

由此可得，可以用物理可实现的分析 CMF，如在第一个例子中，直接显示 XYZ 分量。但是这需要非物理性的显示基色光谱功率分布。或者，可以使用物理性的显示基色光谱功率分布，但是这又需要物理不可实现的 CMF。作为色彩视觉运行的推论，并不存在与恒正分析函数对应的非负显示基色光谱功率分布。

对于这一难题，解决方法是，在图像信号处理过程中构造一个 3×3 矩阵乘法，而不是直接使用图像信号直接驱动显示器。考虑这些显示基色：600nm 的单色红光、550nm 的单色绿光和 470nm 的单色蓝光。式 (26-2) 的 3×3 矩阵可以用来处理 XYZ 值，把它处理成适合于驱动显示的数值。这样的信号处理过程并非只是为了令人更加满意，而是在图像精确再现时必不可少的。

每种彩色摄像机或数码静态相机都需要通过三种不同的光谱特征来感知图像。在数码相机和消费便携式摄像设备中，通常存在一个单独区域阵列电荷耦合器件（CCD，单片

式), 每个 2×2 的阵列内包含三种不同种类的滤光器。通常使用的滤光器呈现红、绿或者蓝三色。其中绿色滤光器复制 2×2 阵列中的两个感光区域。这种方式损失了光和精度。传统的演播室摄像机将入射光用棱镜滤光器进行分光隔离。每个分量都有其专用的 CCD 感知器 (3CCD 或者 3CMOS)。这样一套光系统将不同波长段的光进行分离，并且还不吸收其中的任何光，从而实现了比马赛克传感器高出两个灵敏度的感知特性。

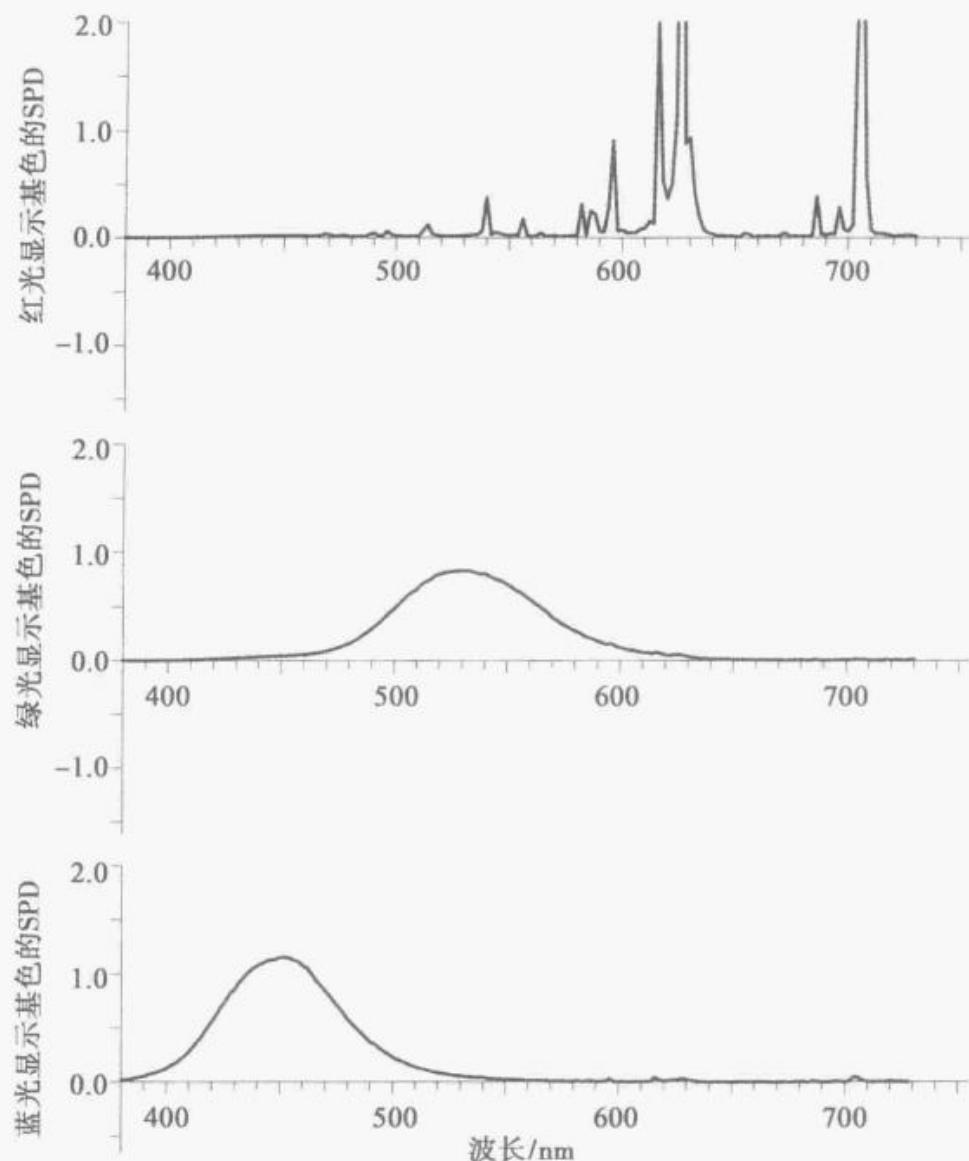


图 26-6 BT.709 显示基色的 SPD

这些光谱功率分布函数具有服从 SMPTE RP145 标准的色度坐标，与 BT.709 标准的类似。许多光谱功率分布 (SPD) 产生相同的色度坐标，这组特定的光谱功率分布 (SPD) 是由索尼单枪三束显示器产生的。与绿光和蓝光使用的荧光粉特性不同，红色基色使用稀土荧光粉来产生非常窄的光谱功率分布。

为了避免像素模糊——到底是由一个构成还是由三个基色构成呢？我建议读者将传感元素称为感光单元。

在单片式照相机中，硬件或者固件对每个感光单元的 RGB 值进行插值。在三片式照相机中，分光滤光片安装在一个或者两个玻璃组件上，在光学工程中，该玻璃组件叫做棱镜。但这里指的不是分光的那种棱镜，而是棱镜滤光器。

图 26-7 给出了一系列的光谱灵敏性函数，这些函数是通过安装在实际高清摄像系统中的分光滤光器实现的。这些函数在整个频谱段的每一个区域都是恒正的，所以这些滤光器都是物理可实现的。然而，如果这些信号直接用于驱动 BT.709 显示基色的显示设备，那么就会产生相当差的颜色再现效果。图 26-8 给出了通过 3×3 矩阵变换得到的相同摄像分析函数。这些转换分量将更精确地重现颜色——这些曲线越接近于图 26-5 所示的 BT.709 的 CMF，则再现的颜色精确度就越高。

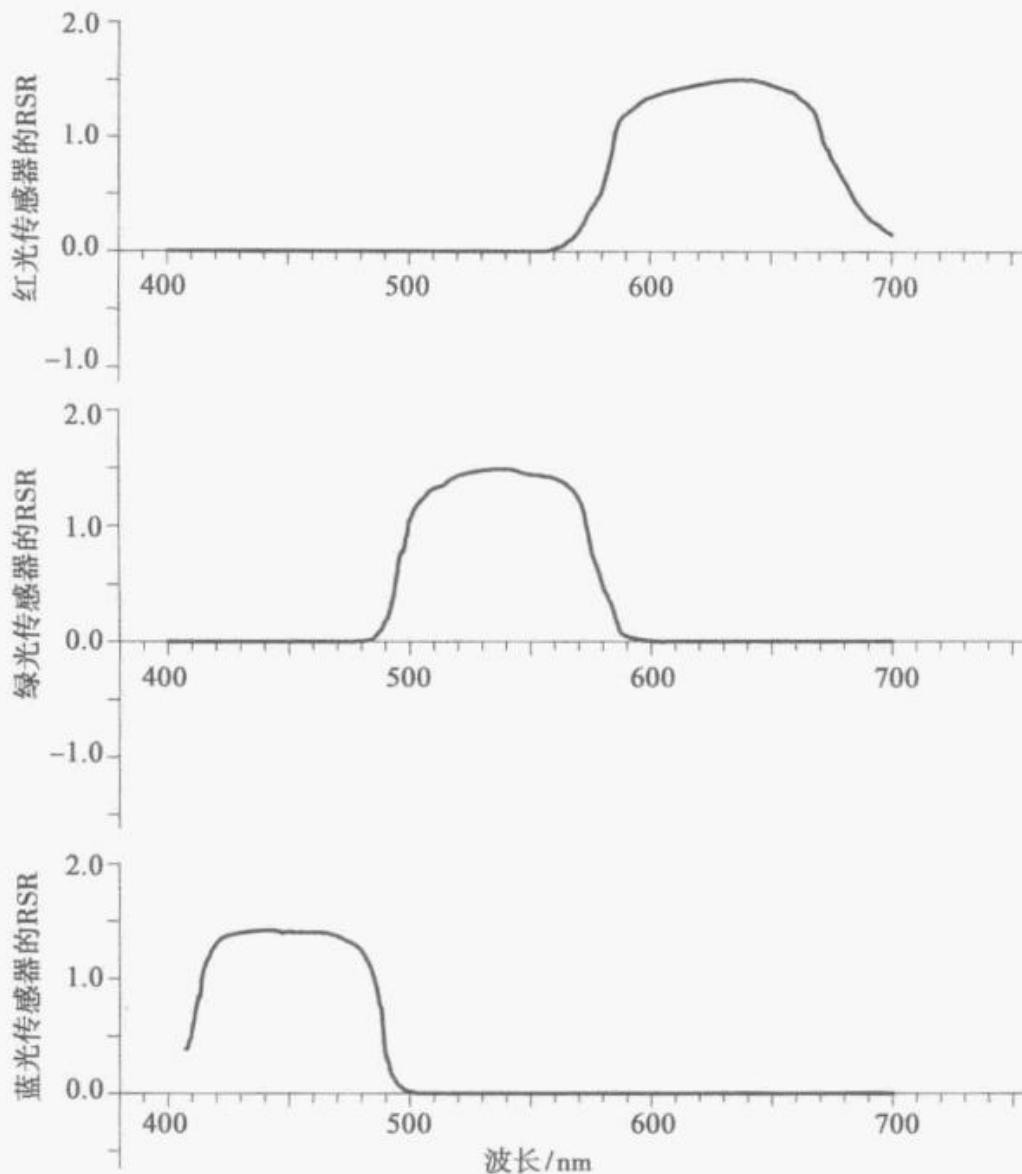


图 26-7 实际相机的相对光谱响应 (Relative Spectral Response, RSR)

这一组光谱响应函数是由 2000-vintage 分光 CCD 视频相机的分光滤光器产生的，我把它称为相对光谱响应

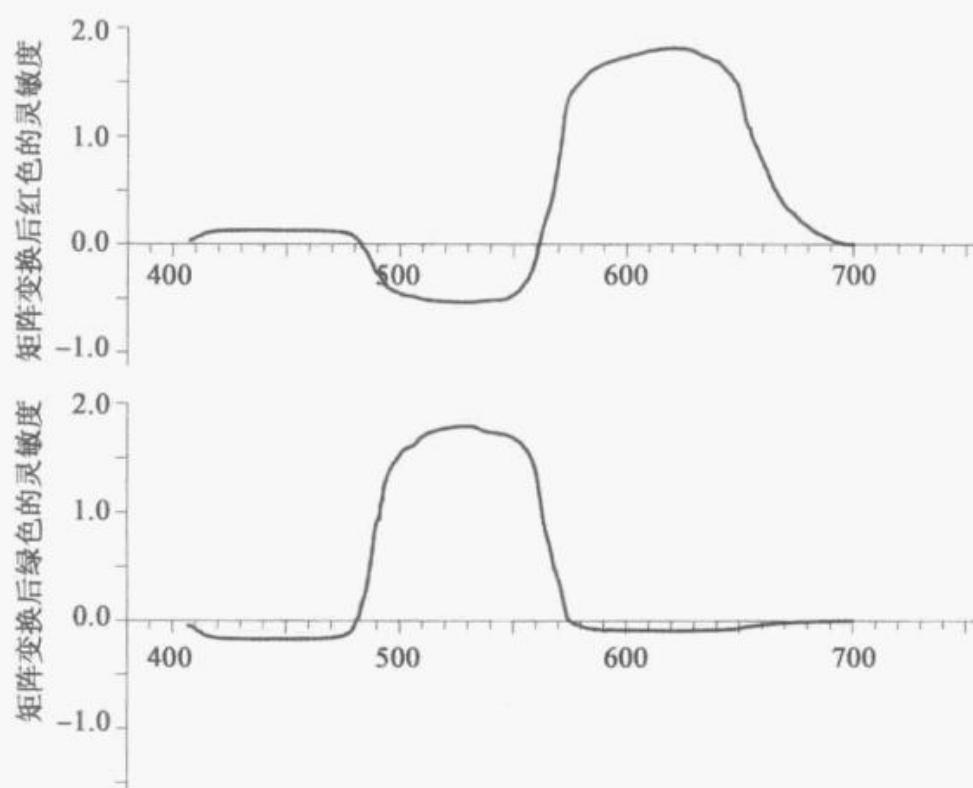


图 26-8 BT. 709 矩阵变换后的有效响应

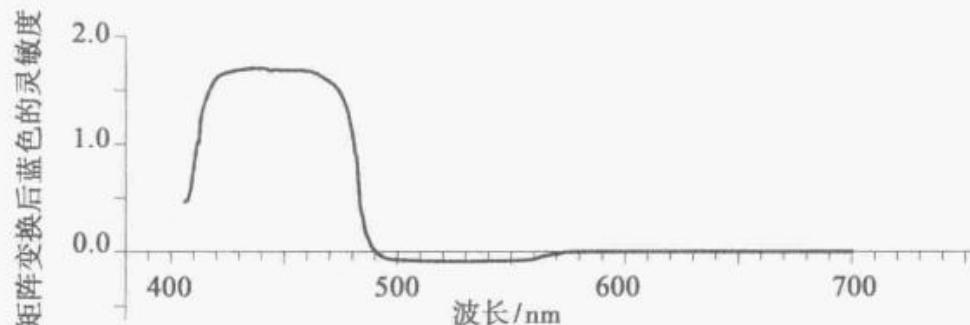


图 26-8 (续)

这些曲线由图 26-7 所示的分析函数使用合适的 3×3 矩阵处理后得到。当这些曲线与图 26-5 所示的 BT. 709 基色相匹配时，该相机感知到的色彩是准确的。

从理论和实际上来说，使用线性矩阵变换来处理摄影图像信号可以正确地捕捉所有颜色，不过，正如本书后续章节将讲到的，在实际中，并不需要捕捉所有颜色。而且，捕捉整个色域上的所有颜色将引入一定噪声，如 26.13 节所述。

26.10 归一化和缩放

$\bar{y}(\lambda)$ 配色函数 (CMF) 通过 CIE 来标准化，因而其最大值处在单位值上。10nm 的 CIE CMF 通常用在图像学中， \bar{y} 曲线积分约为 10.68。对 $\bar{x}(\lambda)$ 和 $\bar{y}(\lambda)$ 函数进行尺度化，以使其积分为相同的值。CIE 通过对它的 1nm 曲线进行插值，派生出了它的 10nm 函数。10nm 的 $\bar{x}(\lambda)$ 和 $\bar{y}(\lambda)$ 函数与 $\bar{y}(\lambda)$ 的积分值并不相同。

CIE 光源被指定在 560nm 处缩放为单位值。其 E 光谱功率分布用一个全 1 的向量来表示。在 CIE 指定的光源和 CMF 缩放中，光源 E 照度值约为 10.68。

在将 CIE 标准应用到成像时，将照度归一化为单位照度是很方便的。光谱由 400~700nm、10nm 为间距的 31 个元素向量组成，也就是说，光强度 E 由元素值为 1/10.68 或者 0.0936 的矩阵组成。

传统情况下，参考白光的照度值一般为 100。

26.11 照度系数

相对照度可以表示成 RGB (线性光) 三刺激值分量的加权和。照度系数可以由 RGB 基色的色度计算得到，这里给出矩阵表示为：

$$\mathbf{C} = \begin{bmatrix} x_R & x_G & x_B \\ y_R & y_G & y_B \\ z_R & z_G & z_B \end{bmatrix} \quad (26-3)$$

任意 XYZ 基色的光谱功率分布的照度系数是 $[0, 1, 0]$ 。任意 X 或者 Z 基色是非物理性存在的，因为含有 0 值照度。

系数 J_R 、 J_G 、 J_B 通过色度和参考白光来计算，公式如下：

$$\begin{bmatrix} J_R \\ J_G \\ J_B \end{bmatrix} = \mathbf{C}^{-1} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} \cdot \frac{1}{y_w} \quad (26-4)$$

对于现行的视频标准中的 D₆₅ 参考白光而言， \mathbf{C}^{-1} 通过与矩阵 $[0.95 \ 1 \ 1.089]$ 相乘得到。

那么，照度可由下式计算：

$$Y = [J_R y_R \quad J_G y_G \quad J_B y_B] \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (26-5)$$

这个计算方式可以扩展为从 [R, G, B] 来计算特定色度的 [X, Y, Z]。首先，计算归一化的基元矩阵 (NPM)，用 \mathbf{T} 表示。NPM 由基色和 [R, G, B] 空间的白光参考点确定：

$$\mathbf{T} = \mathbf{C} \cdot \begin{bmatrix} J_R & 0 & 0 \\ 0 & J_G & 0 \\ 0 & 0 & J_B \end{bmatrix} \quad (26-6)$$

对角矩阵中的元素 J_R 、 J_G 、 J_B 对色度矩阵中对应列的缩放，以及要获得预期的白光色度的基色分布平衡都有影响。然后 CIE 三刺激值 (X, Y, Z) 就可以通过特定的 (R, G, B) 矩阵算出，公式如下：

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathbf{T} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (26-7)$$

正如第 10 章所讲，视频系统通过计算非线性 $R'G'B'$ 分量的加权得到亮度值。虽然这种计算方式会产生非对比性照度误差，但是，使用理论上的相关系数有着很多间接的好处。标清系统的标准相关系数可以由 1953 FCC NTSC 基色和 CIE 的照度 \mathbf{C} 计算得到。高清系统的标准系数可由 BT. 709 基色和 CIE D₆₅ 计算得到。

26. 12 RGB 和 CIE XYZ 之间的转换

一个特定基色集合的 RGB 值可以通过一个 3×3 矩阵变换转换为 CIE XYZ 值。这些变换包括三刺激值变换，也就是，大致与 CIE 配色函数匹配的三种线性光分量的集合。在 XYZ 系统中，颜色用一组恒正的值来表示。SMPTE 已经对这些转换的计算过程进行了标准化。

要从 BT. 709 RGB (参考变光为 D₆₅) 值转换成 CIE XYZ，可以使用如下变换：

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.412\ 453 & 0.357\ 580 & 0.180\ 423 \\ 0.212\ 671 & 0.715\ 160 & 0.072\ 169 \\ 0.019\ 334 & 0.119\ 193 & 0.950\ 227 \end{bmatrix} \cdot \begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} \quad (26-8)$$

该矩阵中间行给出了 BT. 709 标准的照度系数 (虽然 BT. 709 标准指定四个数值)。因为白光已经归一化为单位值，中间行的和值为 1。列向量为纯红、纯绿、纯蓝的 XYZ 三刺激值。要从这样一个矩阵来恢复所有的基色色度，需要对每个 RGB 列向量计算 x 值和 y 值。要恢复参考白点，应先将 RGB 值 [1, 1, 1] 转换为 XYZ 值，再根据式 (25-1) 计算 x 和 y 值。

当针对固定点的计算来构造这样一个矩阵时，需要注意舍入，保证中间行的和为 1。

要将 CIE XYZ 值转换为 BT. 709RGB 值，可以将式 (26-8) 的 3×3 矩阵转换为：

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = \begin{bmatrix} 3.240\ 479 & -1.537\ 150 & -0.498\ 535 \\ -0.969\ 256 & 1.875\ 992 & 0.041\ 556 \\ 0.055\ 648 & -0.204\ 043 & 1.057\ 311 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (26-9)$$

该矩阵含有一些负值系数：超出 BT. 709 色域的 XYZ 颜色值转换为 RGB 分量，其中部分分量为负值或者远大于单位 1。

任意的 RGB 图像数据，或关联 RGB 值和 XYZ 值的任意矩阵，都应当标示期望的

RGB 显示基色的色度值。如果读者遇到了某个矩阵变换或者图像数据，并不包含对于基色色度的参考，就要小心！原作者当时可能都没有意识到，RGB 值必须与色度规范相随，这样才能获得准确的颜色含义。

26.13 矩阵化引入的噪声

虽然位于色度图外部也可能显示颜色，但是在实际中极有可能产生很多问题。现在考虑一个相机，获取 XYZ 三刺激值，然后根据式（26-9）将其转换为 BT. 709RGB 值。在矩阵左上方的系数 3.240479 决定 X 分量到红色分量的转换。一个 4 位数字编码的 X 分量将因 13 位噪声编码的引入而变为红色：在该变换中存在与大相关系数相关联的噪声消除，这一消除在高质量相机设计中有很重要的作用。

26.14 RGB 系统之间的转换

在某个采用一种基色集合的系统的 RGB 值，可以通过一个 3×3 线性光矩阵转换为另一个基色集合。源空间中的 $[R_s, G_s, B_s]$ 三刺激值（用 s 下标来表示）可以转换为目标空间（用 D 下标来表示）中的 $[R_D, G_D, B_D]$ 三刺激值，从对应的色度和参考白光中计算出矩阵 T_s 和矩阵 T_D ：

$$\begin{bmatrix} R_D \\ G_D \\ B_D \end{bmatrix} = T_D^{-1} \cdot T_s \cdot \begin{bmatrix} R_s \\ G_s \\ B_s \end{bmatrix} \quad (26-10)$$

下面的公式主要应用在相同光源系统之间转换。如果光源不同，可能需要某个色度适应性转换（CAT）。在这种情况下，将在矩阵 T_D^{-1} 和矩阵 T_s 之间插入某个合适的 3×3 矩阵（例如 Bradford 转换）。

举一个例子，由 SMPTE RP145 RGB（例如 SMPTE 240M）到 BT. 709 RGB 的转换：

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = \begin{bmatrix} 0.939\ 555 & 0.050\ 173 & 0.010\ 272 \\ 0.017\ 775 & 0.965\ 795 & 0.016\ 430 \\ -0.001\ 622 & -0.004\ 371 & 1.005\ 993 \end{bmatrix} \cdot \begin{bmatrix} R_{145} \\ G_{145} \\ B_{145} \end{bmatrix} \quad (26-11)$$

以下矩阵将 EBU3213 RGB 转换到 BT. 709 RGB：

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = \begin{bmatrix} 1.044\ 036 & -0.044\ 036 & 0 \\ 0 & 1 & 0 \\ 0 & 0.011\ 797 & 0.988\ 203 \end{bmatrix} \cdot \begin{bmatrix} R_{EBU} \\ G_{EBU} \\ B_{EBU} \end{bmatrix} \quad (26-12)$$

从典型索尼单束三枪 RGB 转换为 BT. 709 RGB，可以使用如下变换：

$$\begin{bmatrix} R_{709} \\ G_{709} \\ B_{709} \end{bmatrix} = \begin{bmatrix} 1.068\ 706 & -0.078\ 595 & 0.009\ 890 \\ 0.024\ 110 & 0.960\ 070 & 0.015\ 819 \\ 0.001\ 735 & 0.029\ 748 & 0.968\ 517 \end{bmatrix} \cdot \begin{bmatrix} R_{SONY} \\ G_{SONY} \\ B_{SONY} \end{bmatrix} \quad (26-13)$$

RGB 系统的转换过程可能发生 RGB 分量超出色域的现象，一个或多个 RGB 分量为负值或远大于单位值。

这样的转换只有用到三刺激（线性光）分量值时才能得到准确的结果。在理论上来讲，把非线性的 $R'G'B'$ 从一种基色系统转换为另一种基色系统用转换函数的逆变换来恢复相应的三刺激值，计算矩阵乘法，然后再次使用变换函数。然而，式（26-11），式（26-12），式（26-13）都类似于单位矩阵：对角元素几乎为单位值，非对角元素近似为 0 值。在这种情况下，如果转换关系在非线性 $R'G'B'$ 域计算，那么结果的误差会相对较小。

26.15 相机白色基准

在电视领域，默认认为照相设备在色度为 CIE D₆₅ 的场景中工作。事实上，电视设备通常由温度约 3200K 的钨丝灯点亮。而且场景照度通常定义在光谱的短波长区（蓝光）。这一情况通过白平衡予以补偿——也就是说，调节相机中红、绿、蓝分量的增益，使得漫反射白光对象在场景照度与 CIE D₆₅ 具有相同的三刺激值时会将该值反映出来。在演播室摄影产品中，对于白平衡的控制是可以实现的。在消费类摄影产品中，激活白平衡选项将使设备对图片的红、绿、蓝光进行整合，并且调节它们的增益来实现总和的平衡（这种调节白平衡的方法有时称为灰度综合）。

26.16 显示器白光基准

在加性混合中，重现的图像照度完全由显示设备生成。在实际中，重现的白光由显示器的特性决定，并且不取决于观看显示器的环境。在完全黑暗的观看环境下，如影院，这样的照相设备是可以令人满意的；有较宽的范围的色度可以认为是“白光”。然而，在视场范围内存在不指向显示器的指引标志时，观看者的白色光感很容易被干扰，甚至会由其平常周围的白色光感而代替。为了避免主观误匹配，显示器重现的参考白光色度和周围环境的参考白光应该相当接近。SMPTE 已经对演播室视频中使用的参考白光照度进行了标准化。这些标准指定了重建的参考白光的照度水平为 $120\text{cd} \cdot \text{m}^{-2}$ ，而周围环境的照度为参考白光 10% 的中性灰色。在欧洲，EBU R23 指定参考白光照度为 $80\text{cd} \cdot \text{m}^{-2}$ 。

现代的蓝色 CRT 荧光粉对于人类视觉上的表现力要优于红色和绿色的荧光粉的表现力。直到近年来，计算机显示器的明亮度要比颜色准确性更被重视。在以尽可能小的色彩精度损失为代价来实现小的明亮度增加这一需求的推动下，计算机显示器制造商采用了色温为 9300K 的参考白点，产生的白色中包含的蓝色色度为电视中使用的标准 CIE D₆₅ 参考白色中蓝色色度的 1.3 倍。因而，计算机显示器和计算机图像总会给人一种看起来偏蓝色的感觉。这种情形可以通过将参考白光调整或校正为较低色温来实现。

亚洲的演播室视频标准需要以色温为 9300K 的白光为参考来观看。这种做法显然源于鉴于肤色不同引起的文化上的偏好。

26.17 色域

使用 CIE 分析函数来分析场景将会产生所有颜色的三个不同的分量。但是，将其转换为适合于物理显示器基色的分量时，部分颜色——那些色度坐标处于基色三角形以外的颜色——将会有负的分量值。同时，基色三角形以外的颜色可能有一到两个基色会超过单位值。这些颜色不能够被正确显示。显示器通常会切断负值信号，且将超过单位值的数据饱和化。在色度图上来看的话，基色三角形以外的颜色被重现在三角形的边界上。

如果某个摄像设备设计为采样所有颜色，那么其复杂度必须非常高，且它的性能也明显低于针对小范围颜色捕捉的摄像设备。幸运的是，在自然界和工业界中遇到的颜色只是所有颜色中的一小部分组合。虽然，对于一些仪器来说，例如色量仪，必须测量所有颜色，但是在某个成像系统中，通常只关注那些经常变化的颜色。

M. R. Pointer 定义了经常发生变化的“实表面颜色”。这些自然出现的颜色通常处于色度图的中心位置，在这一区域的颜色可以通过选择合理的物理基色集合来表示。如果某个成像系统能够显示这些颜色中的全部或者大部分，那么该成像系统就相当好了。

BT. 709 表现相当好，不过，传统胶印的许多颜色——特别是那些处于青色色域区的颜色——并不能由全正的 BT. 709 RGB 值来表示，要包含这些颜色就需要宽色域的颜色重现。

26.18 宽色域重现

对于彩色电视经历的大部分历史来说，相机的设计都是基于彩色 CRT 的彩色再现能力的假设进行的。但是现在，视频生成设备通常设计成具有比传统电视广播更宽的应用范围。让数码相机适于在宽范围应用中生成图像，促使适应更宽色域范围的新视频标准的诞生。

xvYCC (“x. v. Color”) 原理将是新一代高清系统更宽色域范围内再现图像的理论基础。该原理利用原有 BT. 709 基色的 RGB 三刺激值，但是其范围扩展到 -0.25 至 +1.33，远远超过了 0 至 1 的范围。这些低于 0 和高于 1 的值使 RGB 值能够表示 BT. 709 基色区域以外的颜色。当扩展 $R'G'B'$ 值矩阵化时，得到的 $Y'C_B C_R$ 值位于有效区域内—— $Y'C_B C_R$ 空间内“合法”RGB 区域以外的立体空间用于转换宽色域颜色。

或许，首个适合于线性光基色低于 0 值和高于单位值的图像编码系统在 LEVINTHAL, ADAM and THOMAS PORTER (1984), “Chap: a SIMD graphics processor,” in Computer Graphics 18 (3): 77-82 中首次出现。

26.19 自由尺度色域、自由尺度指数

最新的 SMPTE 标准支持宽色域图像扩展。FS 代表自由尺度，任意的色度值的图像数据都可以被表述出来，该标准使用标号 $R'_{FS} G'_{FS} B'_{FS}$ 来表示宽色域的颜色组件。标准标题中的“709”表示选择 BT. 709 标准来转换图像数据。FS 的默认值反映的是索尼宽色域 F23、F35、F65 照相机的相关指标。该标准为准指数 OECF 提供了非默认值。

ST2048 中提到“色度坐标”时出现了很多次“三刺激值”。这引起了很多色彩和图像科学家的侧目。

颜色空间通过基色和参考白光的色度坐标与一个参数化定义的非标准 OECF 来共同定义。与上余量和下余量这些典型的非线性区域不同，并没有针对低端和高端空间的相关规定。该标准并不指定如何承载图像数据值，但是，每个分量大概需要使用 10 位的数据。

非标准指数 OECF 通过一组 4 个数字参数和第 5 个参数（曝光率）值 K_{EXT} 来描述。 K_{EXT} 大于 0 代表曝光不足， K_{EXT} 等于 1 代表正确曝光（默认情况）， K_{EXT} 小于 1 代表曝光过量。

该标准定义了色彩 VANC，一个携带色彩元数据的辅助数据包——也就是基色的色度和参考白光、典型指数 OECF 的四个参数、 K_{EXT} 和 12 个与 OECF 的大小参数。据推测，DI 采集将通过色彩 VANC 携带的参数来构建色彩传输。

色彩 VANC 发音为 colour-VEE-ants。相关标准 ST2048-2 建议将色彩 VANC 安置在 1125-线接口 18 线的有效间隔处。

延伸阅读

如果读者需要可读性强的图像编码系统简介，可以参考 DeMarsh 和 Giorgianni 的相关著作。若是需要更简洁，偏向于技术处理的知识，可以参考 Schreiber 的论述。

对于色彩成像技术的很多方面细节信息，可以参考 Kang（虽然有些过时）或者

Sharma 的类似文章。对于计算机图形中的非线性 *RGB* 讨论，可以阅读 Lindbloom 的 SIGGRAPH 论文。

在计算机图形系统中，只要光线进入眼睛，任何基于三刺激值的系统都可以表示颜色。不过，光与物体的相互作用涉及光谱，而不涉及三刺激值。在计算机生成图像 (CGI) 中，计算实际上包含采样的光谱功率分布，即便只使用了三种采样值（本文中的彩色分量）。

Roy Hall 对这些问题进行了讨论。

伽 马 值

在摄影、视频和计算机图形学中，伽马 (γ) 代表一种数值参数，该单值参数用以估计像素点值到三刺激值映射的幂函数的指数。它是一个神秘而令人疑惑的参量，因为它涉及四方面原理：即物理、感知、摄影和视频四方面。本章将解释伽马值与每个原理是怎样相互联系的。对于这些理论和实际的深刻理解将使读者在创建、处理和显示图像时能够得到好的结果。

照度与强度成比例，对于明亮度、光密度、照度、亮度等词汇，可以参考第 3 章。照度和明度的详细内容可以参考第 24 章。

本章主要着眼于使用视频和计算机图形学技术和设备进行电子图像显示。本书主要处理照度描述，或者像摄影者常说的色调比例。实现好的比例重现是实现好的色彩重现的一个重要步骤（关于色彩重现的其他细节在第 26 章已经有所介绍）。

阴极射线管 (CRT) 本质上是非线性的，显示器表面产生的照度是每一个 (R' 、 G' 和 B') 输入电压的非线性函数。从严格的物理学观点来看，相机中伽马校正可以视为对非线性的预补偿，目的是实现相对照度再现的校正。

电光转换函数 (EOCF) 指的是通过信号处理和显示物理学的结合来实现从电信号到光信号的转换的函数。

感知均匀性曾经在 3.3 节中有所介绍。人类视觉对照度的灵敏性是相当不均匀的，视觉的明度感觉几乎是相对照度的指数为的 0.42 的幂函数。如果一幅图像需要以尽可能小的噪声编码来实现每像素有限数据位下的最佳表示，就必须考虑这种非线性。

将 CRT 的非线性（从物理角度）和明度灵敏性（从感知角度）相结合，将发现惊人的一致：CRT 的非线性与人类视觉明度灵敏度的逆变换有着明显的一致。把三刺激值 RGB 编码成伽马修正信号 $R'G'B'$ 将使每个信号分量的灵敏性达到最大。如果伽马校正并非基于物理原因才在 CRT 中使用，那么，就必须从感知的角度对其进行转换。包括 LCD 和 PDP 在内的现代显示技术都没有 CRT 对应的物理效应，但是 CRT 的非线性已经贯穿到信号处理过程中了。

电光转换函数 (EOCF) 指的是在照相机和扫描仪中，将线性光功率与信号编码相联系的传递函数。在视频技术中，有时也称为：光电传递函数，又叫 OETF。

接下来将从以下这些方面讨论在视频技术中伽马校正的若干种应用：CRT 物理知识、视觉感知的非均匀性的知识，摄影学角度的观看条件知识。

27.1 CRT 物理学中的伽马值

CRT 显示器内部的电子枪包含电压输入和光输出的理论相关关系。这种关系称为五等分幂次定律：显示器屏幕所产生的照度原则上与输入电压的 $5/2$ 次幂成正比。照度大致介于电压的二次到三次幂次之间。该幂函数的指数数值用希腊字母 γ （伽马）来表示。传统意义上来说，CRT 显示器大致服从以下幂函数：视频参考 CRT 显示器的伽马值非常接近于 2.4。

图 27-1 所示的是应用于灰度 CRT 电子枪或者彩色 CRT 的红、绿、蓝各电子枪的幂函数示意图。红绿蓝三个通道其实是非常相似的，但是响应并非完全相同。

CRT 显示器电压到照度函数的非线性函数关系，是由于阴极、栅极和电子束之间的静电作用。这种关系也在一定程度上受电子枪的机械结构影响。与通常观点不同的是，CRT 显示器荧光粉本身具有很好的线性，至少在峰值亮度的 8/10 内是线性的。这里用 Y_D 来表示伽马编码的指数。

该函数包含了视频信号与照度之间的非线性关系，图 27-1 给出了增益为 0.9、1.0 和 1.1（增益有时受对比度控制的影响）时的函数曲线。照度约与输入电压的 2.4 次幂成正比。在 CRT 显示系统中，传统意义上的伽马值，是幂函数的指数。图 27-1 中，对比度控制用视频信号（ x 轴上）的增益来表示，这是一种经常使用的方式。然而，根据幂函数的数学特性，照度输出的缩放，将产生相同的效果。

在视频系统中，CRT 典型的解码伽马值 (Y_D) 处于 2.3~2.4 之间。计算机图形学的从业者有时候声称，伽马值远不同于 2.4。然而，这种测量常常忽略了两个重要事实。其一，显示非线性的最大来源是对于明亮度（或者灰度）控制的忽略。在一个伽马值被有效测试之前，这种控制必须进行适当的调整；其二，计算机系统中通常使用查找表 (LUT) 来进行传递函数的影响控制。不同于 2.4 的伽马值通常是由于函数加载到查找表所导致的。例如，据说 2009 年以前版本的 Macintosh 计算机的伽马值约为 1.8。其实，该数值就是 Macintosh 计算机的查找表默认值，并非其本身的显示数值（显示数值是 2.2 到 2.4）。

在 2009 年发布的“苹果计算机的操作系统 10.6（雪豹）”中，苹果公司采用了默认的 2.2 的伽马值：图形子系统中的 $R'G'B'$ 值现在默认解释为 sRGB 标准的 RGB 值。

理解 CRT 显示器的物理原理是理解伽马校正的重要的第一步，但是这并不是全部工作。

27.2 惊人的巧合

第 24 章描述了照度（一种物理参量）和明度（一种感知参量）之间的非线性关系。明度大约是照度的 0.42 次幂。之前的章节讲述了 CRT 的非线性转换函数怎样将电压信号与照度相关联的。下面是令人吃惊的一个巧合。

CRT 信号到照度的转换函数与人眼视觉由照度到明度的转换关系的逆变换十分接近。

在模拟系统中，用电压来表示明度信息，然后用 CRT 幂函数将其转换为照度信息。为了将感知噪声降到最低，使用感知均匀编码。令人吃惊的是，CRT 函数近乎是人类视觉明度灵敏性的逆过程。

CRT 电压是高效的感知均匀编码！在 LCD、PDP 等显示设备中，用信号处理模拟 CRT 的特性。

27.3 视频系统中的伽马校正

在视频系统中，在摄像头中使用伽马校正来达到两个目的：预校正显示 CRT 的非线

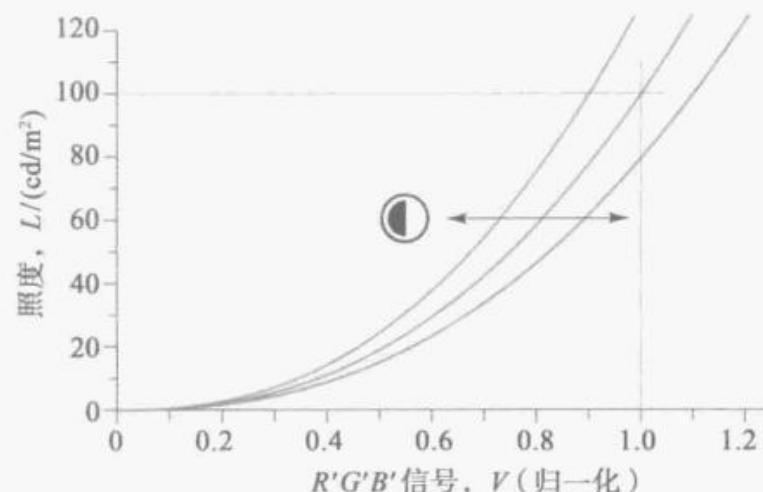


图 27-1 显示电光转换函数

性和在感知均匀空间进行编码。图 27-2 总结了视频系统中图像重现的情景。在左侧，相机上采用伽马校正；右侧的显示器采用伽马校正的逆变换函数。

很多视频工程师对于色彩学并不熟悉。他们只考虑前两个目的，而不在意或者忽略感知均匀编码的重要性。

在电视发展的早期，在感知域进行编码是很重要的，因为需要将空中模拟传输引入的噪声最小化。在模拟录像带的录制过程中，对可见噪声也会进行相同的考虑。当线性光信号量化为有限位的数字信号时，这些考虑也适用于在数字化过程中引入的量化误差。因此，经常对视频信号以伽马校正的形式进行转换。

场景中的相对照度在显示器上呈现出来。不需要再现场景的相对照度水平，所以需要一个适合的缩放系数。然而，场景中检测两种照度水平差从黑到白变化的能力是不均匀的，但是与照度的比例大致呈现一个常数——约为 1.01。在视频中，场景的照度通过类似均方根的函数来转换为一个非线性的、感知均匀的信号。相机设计成模拟人类视觉系统，目的是以人类观看者相同的方式观看场景中的明度。且传输系统引入的噪声对于感知的影响也将达到最小。非线性信号在显示器上再转换为照度。在 CRT 显示器中，固定使用 2.4 次幂函数；在其他的显示技术中，信号处理过程包含了类似的幂函数。

在视频相机中，通常使用非线性传递函数来处理每个 R、G、B 三刺激值，从而对 CRT 的非线性进行预补偿，这个过程称为伽马校正。所使用的转换函数近似为平方根函数。该曲线通常不是精确的幂函数。尽管如此，本书还是将最佳指数表示为编码伽马值 Y_E 。在视频中，伽马校正经常通过相机中的模拟电路（有时也使用数字电路）来实现。在计算机图形学中，伽马校正经常通过把非线性传递函数转换为帧缓冲区的查找表（LUT）来实现。

伽马校正通常依据一个幂函数实现，其形式为 $y = x^a$ （其中 a 是一个常数）。有时，伽马校正被误称为指数函数，其形式为 $y = a^x$ （其中 a 是一个常数），伽马校正与数学中的伽马函数 Γ 无关。

正如第 11 章所讲，基于感知的角度考虑，对于照度大幅低于比原始场景照度的图像，以有限对比度呈现的图像或在昏暗场景下观看的图像，改变其色调是很重要的。昏暗的周围环境是电视观看的特点。在视频技术中，相机是通过将 CRT 的幂函数轻度后补偿实现的，得到一个端到端的幂次为 1.2 的幂函数，如式（27-1）所示。这种后补偿得到的端到端的重建是主观校正（数学上并非线性）。

图像补偿的重要性，以及编码 (Y_E) 和解码 (Y_D) 指数的不同随之产生的不同要求没有被很好地认识到，也没有很好地记录在视频开发文件中。有

$$\begin{aligned} Y_E &\approx 0.5; Y_D \approx 2.4 \\ Y_E \cdot Y_D &\approx 1.2 \end{aligned} \tag{27-1}$$

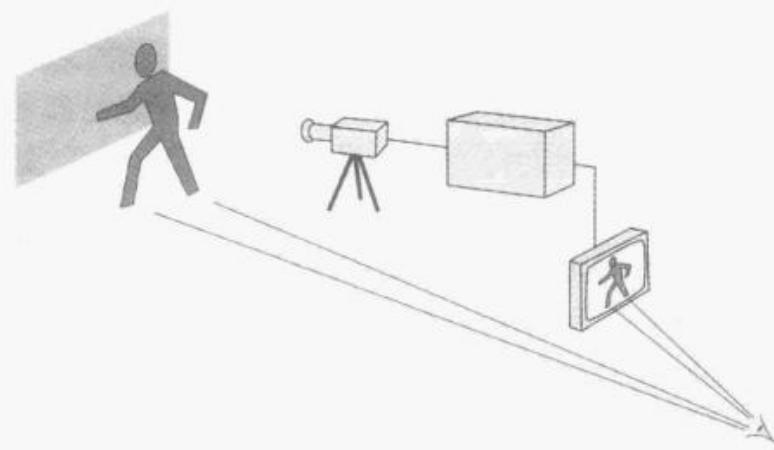


图 27-2 视频中的图像重建

27.4 光电转换函数

很多不同的传递函数已经标准化，并且投入使用。在接下来的几节中，将对这些标准

进行详细讲述。

- BT. 709 标准是指定高清系统基本参数的国际标准。虽然其针对的是高清系统，但是，它对于现行的标清系统也是适用的。而且，这一标准也针对在标清视频系统中的应用做了优化。
- xvYCC 标准延伸出 $Y'C_B C_R$ 和 $Y'P_B P_R$ 而通过编码能适应更宽的色域范围。正如前文所述，该标准没有被使用。
- sRGB 标准指的是通用电子计算机的标准传递函数。
- 1953 年最初的 NTSC 标准制定的传递函数，通常写为 1/2.22，已经被 BT. 1886 标准所取代。
- 欧洲标准针对 576i 的传递函数通常表示为 1/2.8。专业的编码系统从未允许其解码的伽马值达到 2.8。不过，这一值已经被 BT. 1886 标准所取代。

从历史的文件中还不能清楚地看出，经典的 NTSC 2.2 的伽马值和传统的 EBU 2.8 的伽马值是否用于定义相机和显示设备。在娱乐成像时，内容创建者拥有许可证，在获取过程中可对图像数据进行操作，且在后期制作过程中产生预期的画面，可能完全独立于相机中的任何标准的光电转换函数（OECF）。标准的 OECF 规定：OECF 声明了图像数据如何以忠实于内容创建过程的方式来显示。标准相机的 OECF 仅仅作为工程指南来使用。

这里称它为 OECF，是为了与 ISO 14524 协议相一致，在视频学历史上，该协议通常称为光电传递函数（OETF）。

ITU-R 组织建立了 BT. 1361 标准，但是从未使用过该标准。现在，它几乎完全被 xvYCC 标准取代。

27.5 BT. 709 OECF

图 27-3 给出了 BT. 709 国际标准针对高清电视（HD）所定义的传递函数。高清和标清系统中，BT. 709 OECF 作为由场景三刺激值到视频编码的参考映射的标准。它以幂次为 0.45 的纯幂函数为基础。理论上，纯幂函数能满足伽马校正，不过，纯幂函数（幂指数小于 1）的斜率在 0 点处是无穷大。

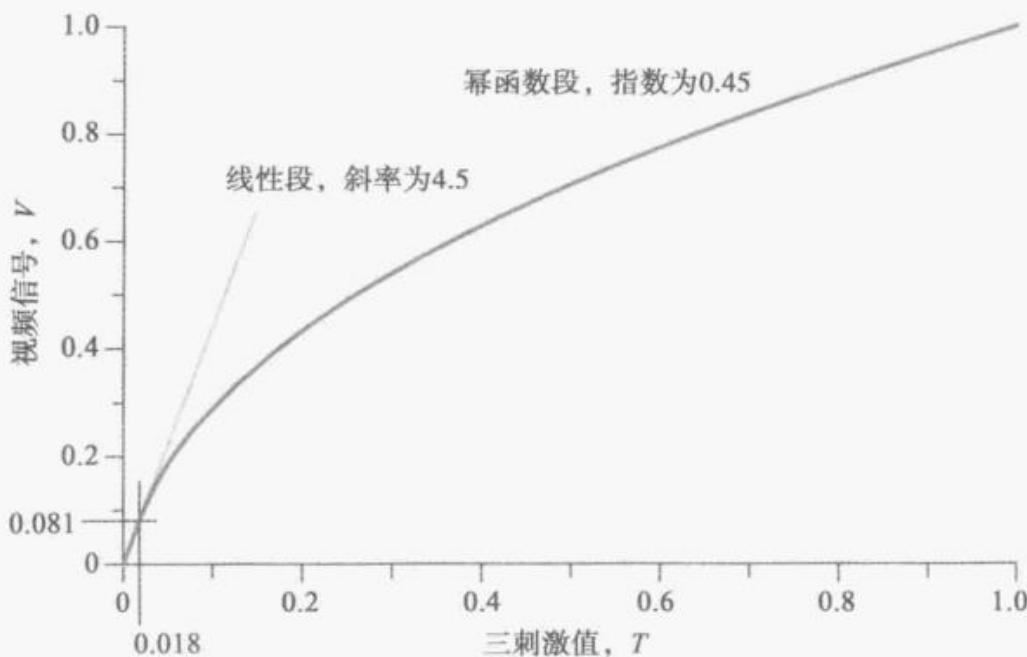


图 27-3 BT. 709 OECF

在诸如视频相机等的实际系统中，为了将图像中的黑色区域中的噪声降到最低，有必要对函数接近于黑色区的斜率（增益）加以限制。BT. 709 指定三刺激值低于 +0.018

时的斜率为 4.5。曲线中的纯幂函数段进行伸缩和偏移以保证该函数及切线在间断点处连续。

BT. 709 编码可以参考下文。三刺激值（线性光）分量表示为 T ，得到的伽马值——用于校正视频信号—— R' , G' , B' 分量之一——用基本符号 V. 709 表示。 R , G , B 值通过指定函数来处理得到 R' , G' , B' :

$$V_{709} = \begin{cases} 4.5T & (0 \leq T \leq 0.018) \\ 1.099T^{0.45} - 0.099, & (0.018 \leq T \leq 1) \end{cases} \quad (27-2)$$

这里的符号 T 代表三刺激值；相同的表达式同样适用于 R , G , B 值。符号 V 代表电压，或者视频或者每像素的编码值，这里没有特意标出。

BT. 709 参考编码公式包含指数为 0.45 的幂函数，我把它称为“标榜”指数。有些人描述 BT. 709 的伽马值为 0.45。广播视频相机的伽马控制的校准值与该值大致相同。不过，缩放因子和偏移量的影响使得整个幂函数与平方根函数非常相似 ($Y_E \approx 0.5$)。有效的幂函数指数，及适于图像补偿校正的值为 0.5。

BT. 709 编码系统假定编码的 $R'G'B'$ 信号将转换为一组三刺激值，其显示时的电光转换函数接近于纯的 2.4 次幂函数。

$$T = V^{2.4} \quad (27-3)$$

指数为 0.5 的幂次通常用在相机中，同时显示设备中指数为 2.4 的设备将产生端到端的指数为 1.2 的幂函数，正如，第 11 章所描述的，它适合于在明亮的环境中获取素材，用于在典型的电视观看环境中显示。在 2011 年，ITU-R 采用了 BT. 1886 标准，该标准指定了应用在高清系统中的 EOCF 为 2.4 次幂函数，参考第 37 章。不幸的是，并没有对参考白光的照度和观看环境进行标准化。

要恢复与场景三刺激值成比例的 RGB 值，假设相机工作在 BT. 709 定义的系数下，也即式 (27-2) 的逆变换：

$$T = \begin{cases} \frac{V_{709}}{4.5} & (0 \leq V_{709} \leq 0.081) \\ \left(\frac{V_{709} + 0.099}{1.099} \right)^{\frac{1}{0.45}} & (0.081 \leq V_{709} \leq 1) \end{cases} \quad (27-4)$$

式 (27-4) 与平方根函数非常相似，不包含图像补偿校正：恢复值与场景的三刺激值是成比例的，并不与显示的三刺激值成比例。由于包含了该表达式，BT. 709 常常引起误解，因为其列出该公式但没有讨论——甚至无人提起图像补偿的重要性。

对于量化为 8~10 位数据分量的详细信息，请参见 4.6 节。

27.6 SMPTE 240M OECF

用于 1125/60、1035i 高清系统的 SMPTE240M 标准比 BT. 709 标准早两年被采纳。事实上，几乎所有在 1988—1998 年生产的设备都采用这些参数。具体情况，请参考本书第一版。SMPTE 240M 指定的 OECF 与显示 EOCF 一起使用，类似于 BT. 1886 中的标准化的做法。

27.7 sRGB 传递函数

sRGB 代表一种用于个人计算、桌面图形和网络图像交换的彩色图像编码的规范。

sRGB 规范指出显示器将会用 EOCF，指数为 2.2 的纯幂函数，对已编码的 $R'G'B'$ 信号进行转换。

与现行的广播演播室实际关联的标准 BT. 709 相比, sRGB 规范对观看的环境亮度水平期望更高。使用 BT. 709 标准编码的图像, 在显示器上显示时其幂指数为 2.2, 对应的端到端幂指数为 1.1, 该值通常低于 BT. 709 标准编码的 1.2 端到端幂指数, 但是, 它更适于高显示照度、高亮度环境和低对比度的典型 sRGB 的显示环境。

sRGB 规范包含由表面定义 OECF 的函数:

$$V_{\text{sRGB}} = \begin{cases} 12.92T & (0 \leq T \leq 0.0031308) \\ 1.055T^{\frac{1}{2.4}} - 0.555 & (0.0031308 \leq T \leq 1) \end{cases} \quad (27-5)$$

该标准未明确是否使用上述函数。显然, 该函数将线性光的值映射为 sRGB 编码, 并且正如读者对 OECF 预期的一样, 它在接近黑色区包含一个线性段。该函数与 BT. 709 OECF 类似。不过, 该过程并没有考虑图片的补偿。总体来说, 正如该标准的 5.1 节所述: 该函数用于描述由呈现在显示器上的三刺激值到 sRGB 编码的映射关系。换句话说, sRGB 值是参考显示的。由 sRGB 指定的编码方式不适合于在图像捕捉的时候进行图像补偿的情况, 例如, 使用数码相机捕捉场景, 这种情况, BT. 709 标准更为合适。

虽然式 (27-5) 包含了幂指数 1/2.4, 代表伽马值为 0.42, 但是, 尺度系数缩放因子和偏移量使整个函数近似于指数为 0.45 的纯幂函数 ($Y_E \approx 0.45$)。因而, 把 sRGB 伽马值描述为 0.42 是一种对读者的误导。

标准规定将 sRGB 分量进行 8 位即 0~255 编码, 不包括上余量和下余量。

图 27-4 画出了 sRGB 的编码函数, 覆盖在 BT. 709 编码函数以及 CIE L^* 编码函数之上。这些编码函数都是近似感知均匀的。不过, 它们之间不能够互换。

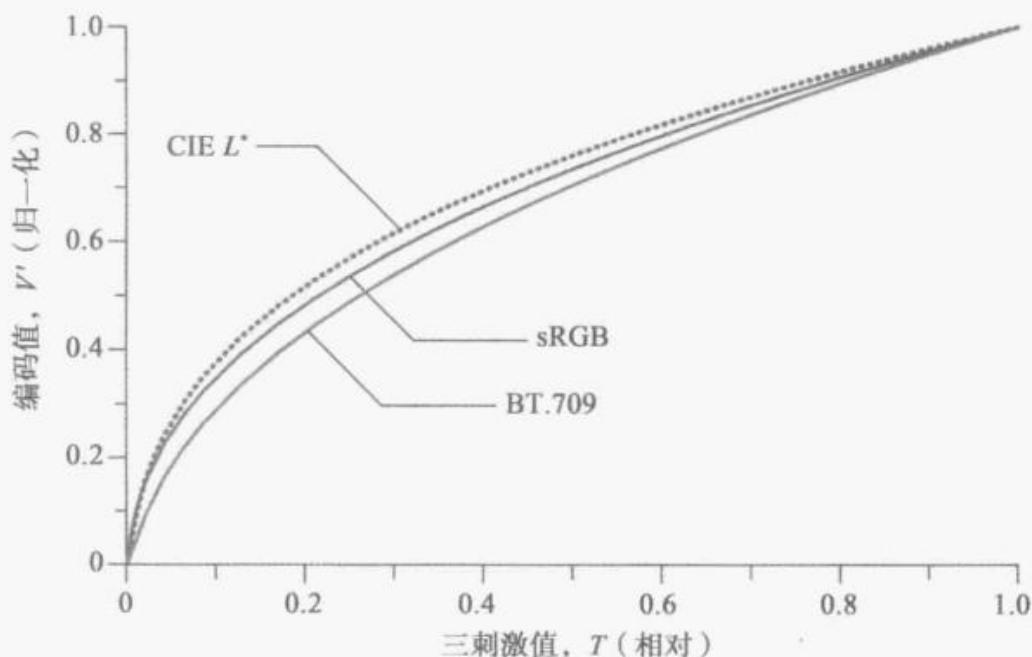


图 27-4 BT. 709、sRGB 和 CIE L^* 的编码函数

27.8 标清的传递函数

在历史上, 高清系统的传递函数几乎从没有被指定过。在 1953 年采用的 FCC NTSC 标准提到: “传输梯度 (伽马指数) 为 2.2”。尚不清楚 2.2 是否能够表示相机的 OECF 或者显示器的 OECF 的特性。对于现代的 CRT 来说, 一般情况下, 幂指数函数都非常接近 2.4。FCC 的声明被广泛解读为编码幂指数应近似为 1/2.2; 而在现代标准, 诸如 BT. 709 中幂指数为 1/2.2 (即近似为 0.45)。不过, 正如本书所提到的, BT. 709 标准的整体有效曲线与平方根曲线非常接近。所以不应该认真对待 FCC 规范: 使用 BT. 709 标准来编码。

576i 的标清标准也没有指明具体的传递函数。在 EBU 规范中，提到了一个指数为 2.8 的假设的显示幂函数。一些研究人员认为这是在暗示编码指数为 1/2.8。然而，这个数值又是高的不切实际。实际上，与世界其他地区相比，欧洲的显示器比较适合使用 BT. 709 标准编码。

令人吃惊的是，并没有现行的标准指定在演播室中的观看环境。只是在 2011 年采用的某个标准中，指明了理想演播室显示器的传递函数。由于缺乏演播室显示器的 EOCD，消费显示设备制造商采用自己的规范（当然是非标准的），这也成为消费显示领域中出现不可预料的图像显示器的原因之一。

在 BT. 601 标准中，用 8 位数据进行编码，不计上余量和下余量，从白到黑区域内约有 220 个等级。对于这个范围内的线性光编码，10 位也足够了。有

$$4.5 \times 220 = 990; 990 \leq 2^{10}$$

$$4.5 \times 880 = 3960; 3960 \leq 2^{12}$$

27.9 位深度的需求

在第 10 章关于相对照度的讨论，即图 10-1 中，本书曾提到，将相对照度直接转换需要大约 11 位数据。该结论是基于两个事实得出的：第一，演播室视频经验证明对于伽马校正 $R'G'B'$ 值转换来说 8 位勉强够——也就是说， 2^8 （即 256）个非线性电平是足够的。第二，用于获取伽马校正 $R'G'B'$ 值的传递函数具有固定的最大斜率。BT. 709 中指定的最大斜率为 4.5。表示线性光需要的编码位数为两个因子的乘积：256 乘以 4.5 得 1152，也即需要 11 位。

在演播室视频中，每个分量用 8 位来表示刚够满足分配需求。如果信号受限于处理操作的话，则需要增加一些空白位用于处理舍入误差。基于这个原因，10 位编码的演播室视频是很常见的。为了在线性光系统中，保持 10 位 BT. 709 的精度，则每个分量需要 12 位。BT. 709 传递函数适用于家用视频显示系统，因为该类系统中，对比度受到周围环境的限制。对于高质量视频，诸如家庭影院，或对于数字影院的高清投影机，则需要更高的最大增益值。当缩放到明度范围的单位值时，CIE L^* 的最大增益值为 9.003；sRGB 的增益值极限为 12.92。对于这些系统，线性光表示比非线性表示的位要多 4 位，非线性表示需要 10 位，也即线性光表示每个分量需要 14 位。

如果 RGB 或者 XYZ 三刺激值分量直接进行转换，那么每个分量值 16 位对于任何实际图像再现系统都足够了。现在线性光 16 位编码已经在高端产品中加以应用了，例如，使用 OpenEXR 编码的场景线性工作流等。不过目前这种方式还没有实时硬件设备。在大多数应用中，都利用了感知的非线性特征，并使用了非线性图像编码。

27.10 现代显示设备中的伽马校正

诸如液晶显示器 (LCD) 等现代显示设备的传递函数与传统的 CRT 显示器有所不同。等离子体显示面板 (PDP) 和数字光处理器 (Digital Light Processor, DLP) 都通过脉冲宽度调制技术 (PWM) 获得表面上的连续显示：它们在本质上是传递函数为直线的线性光设备。线性光设备，例如 PDP 和 DLP，通常都受到描述的“编码 100”问题的影响：在线性光中，为了实现高质量的显示，每分量必须使用超过 8 位的数据量。

PDP 和 DLP 设备通常默认为使用 PWM 技术。然而，其实并不完全是脉冲的宽度进行了调制，只是每周期内的单位脉冲的数量被调制了。

不论传递函数的显示特性如何，以很好匹配感知需求的方式来进行图像编码，从经济

上考虑是很重要的。BT. 1886 EOCF 能够与 CRT 显示器很好匹配，而更重要的是，它与感知特性匹配的也很棒！感知编码的性能优势，采用 BT. 1886 解码设备的广泛使用，以及该标准已经编码的大量可编程素材，这些都使得建立新标准对特定设备进行优化成为不可能。

传递函数与 CRT 显示器不同的显示设备必须进行本地校正，来适应它本身的传递函数到图像交换的标准化传递函数的差异。

27.11 伽马估计

了解到 CRT 显示器是固有的非线性，而且它的响应是基于幂函数的，很多研究人员都曾试图用某个单独的数字参数 γ 来总结 CRT 显示器的非线性度，使用的关系式为：

$$T = V^D \quad (27-6)$$

其中：V 是编码值（或电压值）；T 是照度值（或三刺激值）。

考虑 BT. 601 标准和计算机中常用的全幅度之间的转换，可以参考图 31-3 所示的转换关系。

该模型在任意的伽马值下，都强制零电压值映射到零照度值。由于该模型对零点映射的强制，所以它不能容忍黑电平误差：只有在伽马值估计远小于 2.4 时，取代传递函数上部分的黑电平误差才适用。当伽马估计远大于 2.4 时，取代了曲线下半部分的黑色电平误差才适用——对于一部分低电压区，在零点处饱和。单独的伽马参数能够适应黑色电平变化的唯一方式是改变函数的曲率。该模式下，伽马值的广泛变化导致了伽马值被广为诟病。

更好一些的模型为将幂函数的指数值固定为 2.4，并使用另一个单独的参数 ϵ 来调节黑电平误差，即：

$$T = (V + \epsilon)^{2.4} \quad (27-7)$$

该模型匹配观察的非线性度相比伽马值变化的模型来说要好得多。

一种简单的伽马值估计的技术是使用编码为 0.8 和 0.08 的视频信号的照度度量。再取这两个照度值的以 10 为底的对数，这两个对数值的代数差就是很好的伽马值估计量。这两个视频信号相差 10 倍。0.08 编码的信号对于避免潜在的黑电平误差问题已经足够高；而 0.8 编码值对于避免 CRT 饱和来说也足够低。

图 27-1 画出了几个指数为 2.4 的纯幂函数。这些曲线上每处的伽马值都为 2.4。现在考虑测量 $n+1$ 个视频信号，这些信号在 0~1 之间是以等间距 $1/n$ 分布的（通常使用 10 个值，0.1, 0.2, …, 0.9, 1.0）。用 L_0 来表示由零值信号产生的照度值，用 L_N 来表示由单位值信号产生的照度，平均伽马值估计值为：

$$\gamma = \frac{1}{9} \sum_{i=1}^9 \frac{\lg \frac{L\left(\frac{i}{10}\right)}{L(1)}}{\lg \frac{i}{10}} \quad (27-8)$$

式 (27-8) 对数的底数选为 10。不过，由于要求的对数的比值，与底数无关。该式计算中，照度值 L_1 到 L_N 值必须比 L_0 大。如果视频信号值不是相同间隔，则用 V_i 代替分子中的 i/n ，这里的 V_i 值为适合的视频信号电平，范围严格处于 0~1 之间。

在 EBU. 3325 中描述了该公式的一种变体。该式一般用于家庭影院校准。在我看来，这种计算平均伽马值的公式给出了由参考白光视频级产生的照度，会对估计的伽马值产生过度的影响。如果参考白照度被压低——就像 CRT 进入饱和区时，或者用 LCD 模仿这种情况时一样——所有相关点的伽马值都将变低，平均伽马估计值也会降低。在一些公式中

(如欧洲广播联盟的公式中), 分子中的 L_i 通常用 $L_i - L_0$ 来代替, 即减去零值编码照度, 这里的减法会引入一定误差。

作者认为, 伽马值特征化的一种比较好的方式是, 对合适的近似模型进行数值匹配, 例如 Berns、Mottta、Gorzynski 设计的 GOGO 模型。

历史上通过使用 PLUGE 测试信号, 调节黑电平水平 (亮度或者偏移量), 将 0% 到 2% 的视频信号调整到稍低于可视门限等方法使视频显示器的性能接近于演播室效果。这一处理过程使得零视频信号产生少量的光, 相对于 100nt 的参考白光而言, 其大小通常为 0.01~0.1nt。对于指数为 2.4 的幂函数而言, 视频信号对于绝对的、理论黑色约为参考白光信号的 2%, 也即 10 位接口编码 32。

如果需要确定显示设备的非线性特性, 可以参考 Cowan 的经典文章。除了描述了如何测试非线性度外, 他也描述了如何确定显示器的其他特性 (如白光参考点和其基色的色度), 这些对于准确的颜色再现是很重要的。

27.12 视频、CGI 和 Macintosh 中的伽马值

视频 (个人计算机)、CGI (计算机生成的图形)、Macintosh 中的传递函数在图 27-5 中给出。三刺激值信号从左端输入, 每一行表示了四个函数块, 从左到右, 它们分别是相机或扫描仪的查找表 (LUT)、图像存储设备、输出查找表以及显示器。

在视频中, 图 27-5 最上边一行所示的为应用于相机中模拟视觉 (伽马校正) 的传递函数。信号保留在感知域中, 直到编码在显示器中进行转换为止。在计算机图形中, 像图 27-5 第二行所示, 在线性光域中计算, 然后在帧缓冲区输出的查找表中进行伽马校正。苹果计算机, 如在图 27-5 最下边一行所示, 采用的是一种混合的方法: 扫描仪使用指数为 1/1.66 的幂函数, 查找表中使用指数 1/1.45 的幂函数。使用 $\gamma_E \approx 1/1.66$ 适于预生成的图像, 可以产生一个端到端的指数 1.0。端到端幂函数指数或者图片补偿, 在每一行的最右边用数字给出。该数值是整个系统指数的乘积。有人将其称为系统伽马值。但是这个词引起很多误解, 所以我不使用这样的说法。

在视频技术中, 也就是图中的最上边一行, 相机使用传递函数来完成伽马校正。随后信号保持在系统的感知域, 直到在显示器端转换为三刺激色值。这里用一条斜线来表示使数据不变的输出查找表: 传统的视频系统并不使用查找表, 但是如果用相同的方块来描绘这四行的话, 则比较起来就很清楚。

正如本书第 1 章所讲述的, 计算机图像硬件通常在帧存储以外实现查找表结构。然而, 大多数的计算机软件都适应了没有查找表的计算机显示硬件。当没有查找表时, 编码值直接映射为电压值, 视频也是同样的情况。所以, 图中最上边一行适用于个人计算机。

计算机图形系统通常在帧缓存区存储三刺激值, 并在显示的通道中使用硬件查找表, 来实现伽马校正, 如图 27-5 第二行所示。通常来说, 在输出查找表中加载一个指数为 1/2.2 的幂函数, 可实现系数为 1.1 的图像补偿。

历史上, 苹果计算机的输出查找表使用指数为 1/1.45 的幂函数。John Knoll 的伽马控制面板经常用来加载输出查找表。当伽马值设置为 g 时, 控制面板会将一个指数为 $2.61/g$ 的幂函数加载到查找表中。奇怪的是, 苹果计算机的伽马值先用作指数, 然后再送进缓冲器中 (而在其他计算机中, 加载在查找表中的是幂指数)。因此, 苹果计算机的默认伽马值为 1.8 而不是 1.45。显示伽马值为 2.2 更加合理, 使得三刺激值与编码的比值提高到 1.66 次幂 (见图 27-6)。

在输出查找表中加载斜率变化, 苹果计算机可以通过设置来处理视频 (或计算机) 的

$R'G'B'$ 值。使用 Knoll 的伽马控制面板，可以通过将伽马值设置为 2.61 来实现。

早先由苹果计算机生成的 JPEG/JFIF 文件，其表示 RGB 的显示三刺激值的幂指数提高为 0.6（也就是大约 $1/1.65$ ）。对于苹果操作系统 10.6 而言，苹果软件已经与 sRGB 的显示特性相一致了。

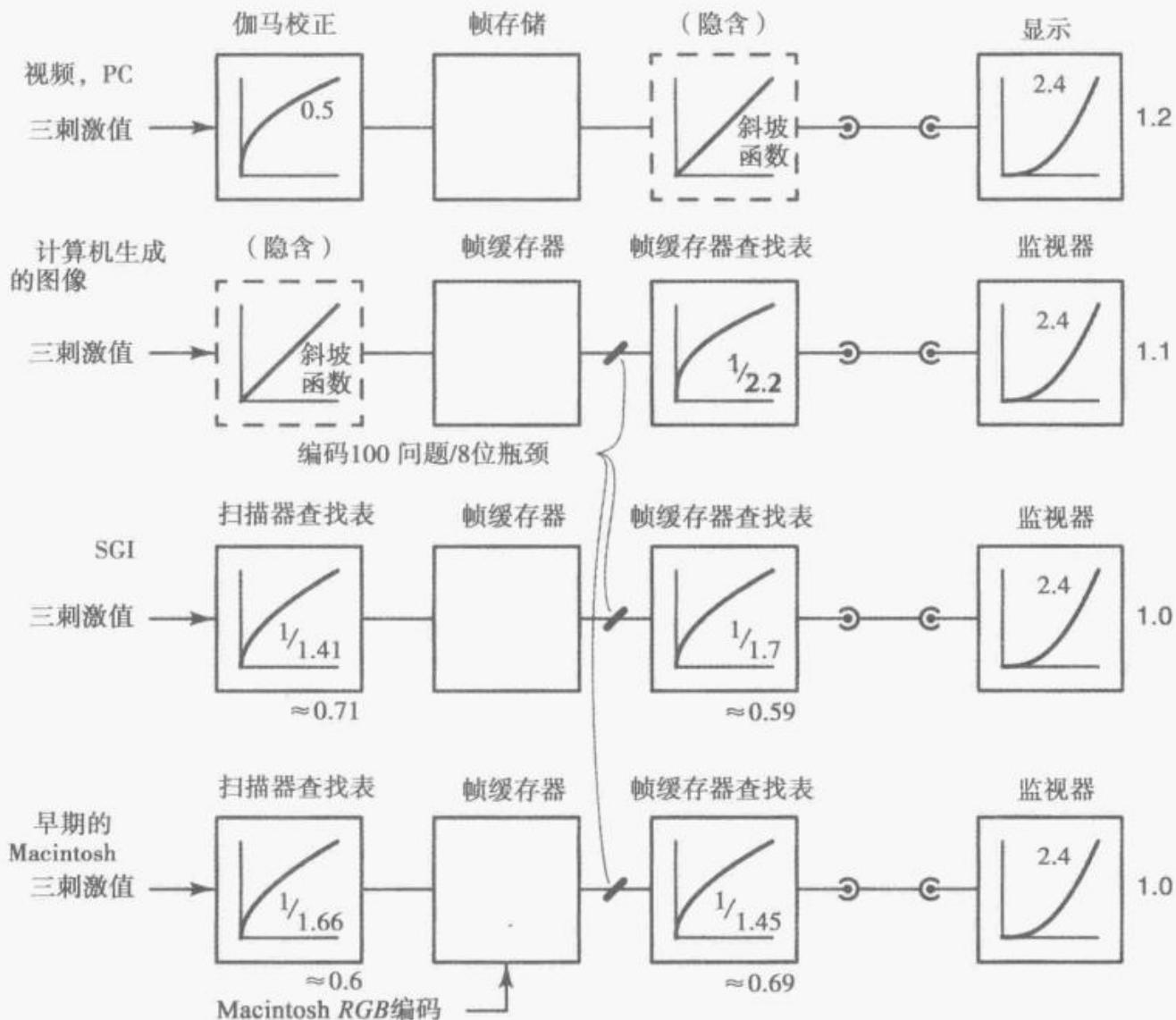


图 27-5 视频、计算机图形图像和苹果计算机中的伽马值

如果桌面扫描设备想要产生正确显示相对照度的 Macintosh 中的 $R'G'B'$ 值，则扫描仪查找表中就必须加载指数为 1.66 的幂函数。在典型的 Macintosh 的操作环境中，指数 1/1.66、1/1.45 和 2.4 结合起来实现一个端到端的单位指数。这种情况适用于扫描照片或者胶印印刷品，这里图像补偿已经融入到图像中了。

对于苹果计算机应用软件产生的 $R'G'B'$ 值而言，应用软件必定会影响部分苹果计算机的伽马校正值，且优先表示苹果图形子系统的 $R'G'B'$ 值，其余部分的值将在输出查找表中实现。扫描时，部分伽马校正也受到扫描仪中查找表的影响，其余部分在输出查找表中实现。

半色调打印由于点增益现象的存在，因而具有默认的非线性。来自于打印页的反射系数约与 1.8 次幂函数 [1-CMYK] 编码值成正比。苹果系统的 $R'G'B'$ 值不是感知均匀的。然而，很巧合的是，苹果系统的 $R'G'B'$ 编码几乎与半色调打印的点增益完美匹配。这使得苹果计算机在图像学领域具有统治地位，也促成 1.8 伽马值编码系统成为图形学的事实标准。

在图 27-5 所示的每一行的右侧，已经用黑体标出了通常使用的补偿值。图 27-5 也给

出了一种用于视频应用中的端到端的 1.2 次幂函数。对于计算机生成的图像，本书给出了典型值 1.1。对于苹果计算机，也描绘了扫描预补偿图像时的常见情形。这时，端到端的幂函数指数为单位值 1。

正确的计算机图像显示依赖于图像子系统输出端的传递函数。如果由计算机产生的图像经过苹果 10.6 之前版本系统查找表 (LUT) 的经典 $1/1.45$ 次幂函数和 2.4 次幂函数的显示器显示，则中段色调会显示得过于明亮：128 编码会产生比原指标高 1.5 倍的照度。反之，如果某个图像产生于早期的低于 10.6 版本的苹果计算机系统（这些系统采用的是 $1/1.45$ 次幂函数），但是却在没有该传递函数的计算机上显示，则中段色调将会变得十分暗淡。通用计算机和苹果计算机默认的 $R'G'B'$ 编码值和显示照度系数之间的关系如图 27-6 所示。通用计算机和苹果计算机的伽马值不同，是由于显示系统 $R'G'B'$ 编码的插值算法不同。在通用计算机中，输出查找表 (LUT) 要么不存在，要么编程时没有考虑到，而编码值服从显示器端的 2.4 次幂函数（正如底端曲线所示）。在苹果 10.6 版本之前的系统中，默认的输出查找表在编码值中引入一个指数 $1/1.45$ 的幂函数，然后显示器使用常用的 2.4 次幂函数。这两个函数的综合影响产生指数为 1.66 的幂函数，正如图 27-6 上方曲线所示，该函数将苹果计算机的编码值与显示的相对照度联系起来。Macintosh 计算机 Mac OS X 10.6 之前的版本，采用的是图 27-6 中最下边一行的技术。默认情况下，输出查找表加载一个指数为 $1/1.45$ 的幂函数。默认查找表和常用的 2.4 次幂显示函数的结合得到指数为 1.66 的幂函数，该幂函数将 Macintosh 中的 $R'G'B'$ 值（例如在数据结构或者 PICT 文件中存储的数值）与显示三刺激值相关联起来。

伽马值偏移是指，某种意料之外的有效解码伽马值的变换，这种变换来自于苹果计算机相关的伽马校正在苹果计算机的视频输入和输出中的不经意使用。伽马值偏移通常包含无意中使用了 1.45 次幂函数或者 1.45 的倒数 0.69 次幂函数。

27.13 计算机图形学中的伽马值

计算机图像生成软件系统通常使用三刺激值的近似来执行亮度、阴影、景深和抗混叠的相关计算，以便实现光线的物理融合。在帧缓存中存储的数值通过显示前的硬件查找表来实现。如果线性光数值存储在帧缓存中，则查找表就可以实现伽马校正。CRT 显示器中的幂函数以伽马校正信号电压的形式在屏幕表面显示已校正的照度值。软件系统通常提供默认的伽马值和一些改变默认值的方法。

BT. 709 函数很适合于高亮度水平（2000lx 甚至更高）的图像数据的生成，这些图像用于在约 100nt 的昏暗环境中使用。对于其他的生成和观看环境的讨论，请参考第 11 章。

帧缓存查找表使软件执行一些小技巧来改变图像数据的外观，而不是从根本上改变图像数据。要允许用户使用如精确再现图像等的特性，应用程序应该以结构化的方式访问图形系统提供的查找表，而不是直接对查找表进行操作。

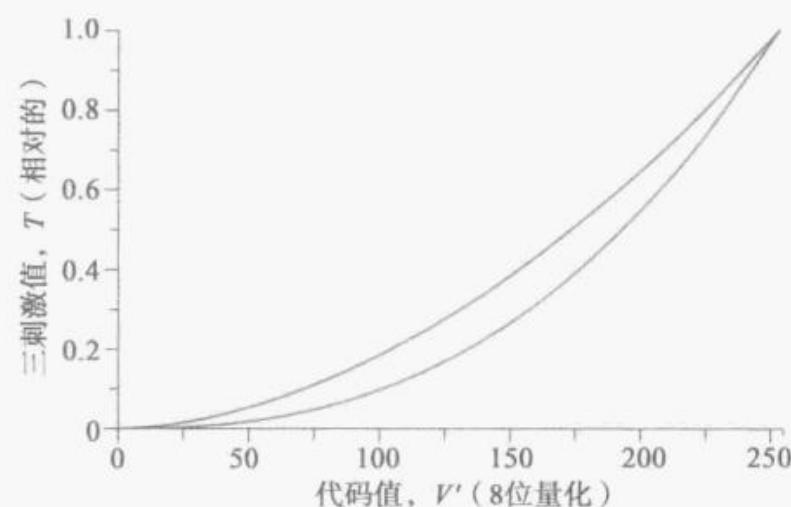


图 27-6 通用计算机与经典苹果计算机的伽马值

27.14 伪彩色中的伽马值

第6章描述了伪色彩系统中彩色查找表如何存储直接映射到显示器的电压值。对于伪彩色应用程序来说，向图像系统提供 $R'G'B'$ 彩色编码值是很便利的，这些编码值已经针对典型的监视器和典型的观看环境进行了伽马校正。

文件中存储的伪彩色图像都有相随的彩色映射图，映射图中的 $R'G'B'$ 编码值服从电光转换函数（EOCF）形式，在显示器端近似为指数为 2.4 的幂函数。

27.15 8位线性编码的局限

正如 27.13 节所提到的，补偿合成图像的计算机图形系统，通常在线性光域或者粗略的形容为“强度域”中进行运算。低端图形加速器历来都是在线性光域中执行 Gouraud 明暗处理，在帧缓存器中存储 8 位分量值的。3.4 节解释了线性光表示方法，每个分量只用 8 位的数据表示是无法获得高质量图像的，通常这样的图像仅仅能够显示大致轮廓。由感知效应加强轮廓的可视性称为“马赫带”。因此，这种轮廓伪影通常称为带状伪影。

高端的计算机图形生成系统（CGI）通常运行于线性光域（伽马值为 1.0），每个分量使用超过 8 位的数据（通常为浮点型数据）。一些系统在软件中进行伽马校正，然后将伽马校正值写入深度有限的帧缓存器中。其他系统采用“深度”缓冲器（数据宽度大于 8 位，通常为 16 位），并在帧缓冲器的查找表中加载单位斜率值。这样的安排最大限度地提高了感知性能，且产生的补偿图像没有 8 位线性光编码产生的量化失真。

27.16 CGI 中的线性和非线性编码

计算机图形标准没有明确提及传递函数，通常隐含为线性光编码。不过，虽然 JPEG 标准也并未提到传递函数，但是隐含的是非线性编码：当 JPEG 系统使用线性光数据时将会得到难以接受的结果。所有的这些标准都使用 RGB 数值进行处理，你可能会认为这些 RGB 数值是可比的，但是事实上它们不是！

图 27-7 举出了两种显示同样 RGB 值 [128, 128, 128] 的系统。左侧显示的是光度计读取出线性场景系统的照度显示值；右侧给出光度计读取到视频系统显示的照度值。在线性场景中，显示照度值为物理尺度的一半，即相对照度值为 0.5。在视频的例子中，显示照度为感知大小的一半，相对照度约为 0.18。两个例子中的数值都用 RGB 表示，不过，这些数值并不具有可比性，这一差异体现了图像文件变换中的一个重要问题。很多图形图像文件并不包含任何传递函数的相关信息。如果要对 RGB 图像进行转换而不指定传递函数，那么图像显示的时候将产生巨大的差异。

数字图像处理领域中很少对线性编码和非线性编码加以区分。此外，如果提到了强

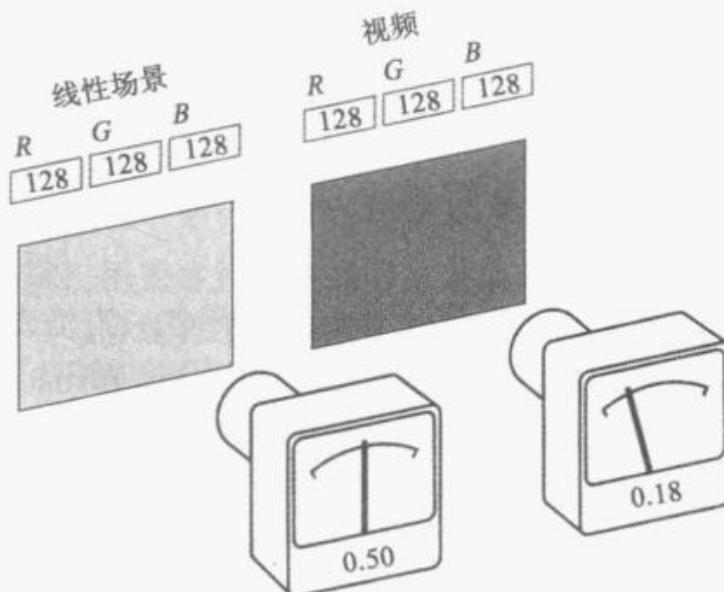


图 27-7 成像标准中的线性和非线性编码

度这个词，则要小心：图像数据可能是以线性光的形式来表示的，与强度成正比。不过，像素分量值通常与某个小区域的传感器或者显示器相关，因而，它的单位应该为每平方米物理量的值，所以辐射度、照度、相对照度和三刺激值，技术上来说都是正确的。所有这些数值都是与强度成正比的，但是它们并不具备强度的单位，并且也不用强度值来描述。

使用 JPEG 交换格式 (JFIF) 的文件就称为 JPEG 文件。JFIF 规范 1.02 版本指定了线性光编码（伽马值为 1.0）的状态。这在实际中是很少使用的。通常情况是，图像数据用 2.2 次幂 EOCF 进行编码，参见 27.12 节。

第 28 章

亮度与色差

本章介绍用于传输源于加性 RGB 基色图像数据的彩色编码系统。文中概述了非线性 $R'G'B'$ ；解释了作为这些非线性信号加权和的亮度的形成，记为 Y' ；介绍了色差（色度）分量 $[B'-Y', R'-Y']$ 、 $[C_B, C_R]$ 和 $[P_B, P_R]$ 。

视频编码系统的设计必然是植根于对人类色彩感知的具体认识。然而，一旦这些知识体现在编码系统中，之后要做的就是物理、数学和信号处理。本章仅涉及后面这些领域。

28.1 色敏度

如前文所述，单色视频系统能理想地感知相对照度。如“视频中的伽马校正”所述，照度经相机的伽马校正电路变换为一种考虑到亮度感知特性的信号。在接收机中，显示器——历史上是 CRT——本身施加了必需的逆传递函数。

如 26.1 节所述，一幅彩色图像是以红、绿、蓝三个分量进行感测的。为了最小化噪声或量化的可见度，RGB 分量应进行非线性编码。

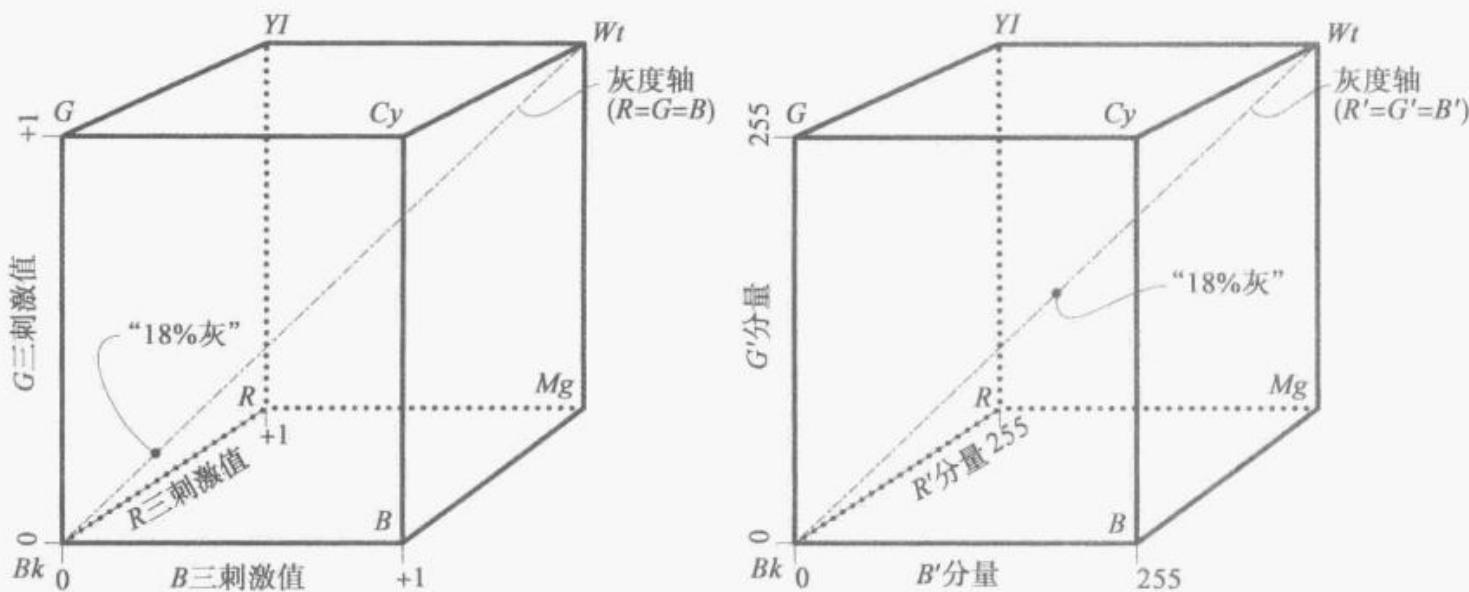
28.2 RGB 和 $R'G'B'$ 彩色立方体

如第 26 章所详细描述的，红、绿、蓝三刺激值（线性光）基本分量可看作是在一个三维颜色空间中的坐标。在 0 和归一值之间的坐标值定义了空间中的单位立方体，如图 28-1 的左图所示。线性光编码用于 CGI，以模拟物理光。然而，正如 27.3 节讲到的，8 位线性光编码的感知性能差：每个分量需用 12 位或 14 位才能达到优秀的质量。通过使用模仿人类视觉的非线性亮度响应的非线性编码，有限的位数就能构成最佳的感知利用。如第 27 章的内容，在视频、JPEG、MPEG、计算机、数字摄影和许多其他的领域中，一个非线性传递函数应用于 RGB 三刺激值信号，生成非线性编码的（经伽马校正的）分量，用带撇号的符号 $R'G'B'$ 标记。用类似 BT.709 或 sRGB 的传递函数进行 10 位非线性编码，可获得优秀的图像质量。

在 PC 绘图中，8 位非线性编码很常见，在图 4-1 所示的量化器传递函数之后， R' 、 G' 、 B' 的范围是 0~255，得到的 $R'G'B'$ 立方体如图 28-1 的右图所示，总共可表示 2^{24} 种，即 16 777 216 种颜色。不是所有这些颜色都能在视觉上进行区分，也不是所有这些颜色都能被感知，但它们还都是颜色。如 4.6 节所述，演播室视频使用上余量和下余量：8 位 $R'G'B'$ 在黑色和白色之间有 219 种编码，总计有 220^3 或 10 648 000 种码字。

在视频中，码字（或码点）指的是三个整数值的组合，如 $[R', G', B']$ 或 $[Y', C_B, C_R]$

传输图像的 $R'G'B'$ 分量的缺点是，每个分量都需要相对高的空间分辨率：传输或存储用 $R'G'B'$ 分量表示的彩色图像需要 3 倍于灰度图像的容量。人类视觉对彩色信息的空间分辨力比对亮度的空间分辨力低得多。由于视觉的彩色分辨力差，彩色图像可编码为一个表示亮度的宽带单色分量和两个携带彩色信息的窄带分量，每个窄带分量的空间分辨率比亮度的低很多。在模拟视频中，每个彩色通道的带宽一般是单色通道的 $1/3$ 。在数字视频

图 28-1 RGB 和 $R'G'B'$ 立方体

中，每个彩色通道的数据传输速率（或数据容量）是单色通道的一半或更少。存在有力的证据说明，人类视觉系统在视网膜上形成了一个非彩色通道和两个彩色的色差通道。

绿色主导照度：60%至70%的照度由绿色信息构成。如果另外两个色差分量上的彩色信号选为蓝色和红色，这可使信噪比最大。最简单的从蓝色和红色分量中“去除”亮度的方式是减掉亮度，从而形成一对色差（或不严格地称为色度）分量。

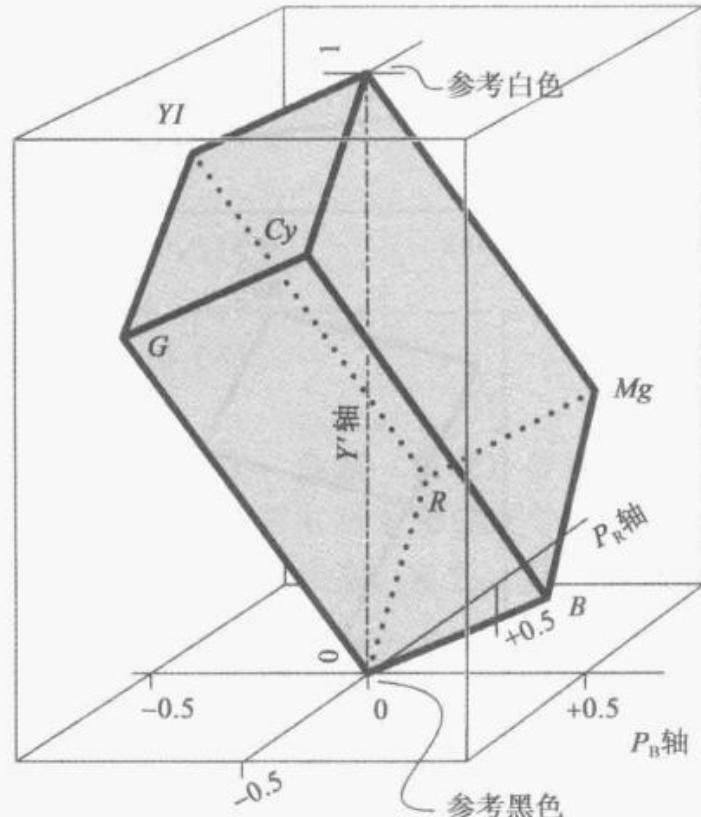
这里的术语色差指的是由两个伽马校正后的彩色分量的差值形成的信号。在其他上下文中，该术语可能指一个两种颜色之间感知距离的数值量度。

在彩色视频中的单色分量可以基于色彩学中的照度（ R 、 G 、 B 的加权和）。然而，如第 10 章的解释，亮度是 R' 、 G' 、 B' 的加权和，所用的系数与用于计算照度的系数相似或相同。用抽象术语来说，亮度范围是从 0 到 1。色差分量 $B'-Y'$ 和 $R'-Y'$ 是双极性的，各自的范围接近 ± 1 。

在分量模拟视频中， $B'-Y'$ 和 $R'-Y'$ 经过缩放形成 P_B 和 P_R 分量。用抽象术语来说，它们的范围为 ± 0.5 。图 28-2 表示单位 $R'G'B'$ 立方体变换为亮度 $[Y', P_B, P_R]$ 。（在用的有不同的接口标准。）在分量数字视频中， $B'-Y'$ 和 $R'-Y'$ 经过缩放形成 C_B 和 C_R 分量。在进行接口偏置前的 8 位 $Y'C_B C_R$ 中，图 28-2 所示的亮度轴应放大 219 倍，色度轴放大 112 倍。

色差信号一旦形成，便可通过亚采样来减小带宽或数据容量，而不会被观察者注意到。这部分内容将在“再论色度亚采样”中介绍。

从图 28-2 能明显看出，当 $R'G'B'$ 信号变换到模拟视频的 $Y'P_BP_R$ 空间时， $R'G'B'$ 单位立方体仅占 $Y'P_BP_R$ 单位立方体的部分体积：仅有 $1/4$ 的 $Y'P_BP_R$ 体积对应所有从 0 到 1 的 $R'G'B'$ 值。因此， $Y'P_BP_R$ 的信噪比相较于 $R'G'B'$ 有损失。不过，能够进行亚采样的优

图 28-2 $Y'P_BP_R$ 立方体

点抵消了这个缺点。

在合法的信号中，各分量不能超过其参考行程。合法的 $R'G'B'$ 信号组合称为有效的。在 $Y'P_B P_R$ 单位立方体之内的信号是合法的 $Y'P_B P_R$ 。然而，这些组合中大约有 $3/4$ 对应 $R'G'B'$ 单位立方体之外的 $R'G'B'$ 组合：虽然合法，但是这些 $Y'P_B P_R$ 组合是无效的，也就是说，它们是非法的 $R'G'B'$ 。

在数字视频中，我们称码字而不是组合。与 8 位演播室 $R'G'B'$ 的 1060 万个码字相比，8 位 $Y'P_B P_R$ 有约 275 万个有效的码字。如果 $R'G'B'$ 转换为 8 位 $Y'P_B P_R$ ，然后再转换回 $R'G'B'$ ，得到的 $R'G'B'$ 将不超过 275 万种颜色。

$$\frac{\frac{1}{4} \times 220 \times 225^2}{220^3} = \frac{2\ 784\ 375}{10\ 648\ 000} \approx 0.261$$

在图 28-2 中， $Y'P_B P_R$ 立方体的绘制是偏离坐标轴的。图 28-3 表示 $R'G'B'$ 棱柱在 $Y'P_B P_R$ 空间中的三个正交视图。亮度轴标记为 Y' ，范围从 0 到 1。色度轴标注了 $[B'-Y', R'-Y']$ 缩放（两个分量的范围分别是 ± 0.886 和 ± 0.701 ）和 $P_B P_R$ （两个分量的范围都是 ± 0.5 ）缩放。显然 $Y'P_B P_R$ 空间的体积范围超出了 $R'G'B'$ 棱柱。新兴的 xvYCC 系统用到在单位 $R'G'B'$ 棱柱外的 $Y'P_B P_R$ 码字，也就是前面提到的“无效”码字，来传送宽色域的颜色。

$C_B C_R$ 分量与 $P_B P_R$ 分量类似，但是码字取值范围是 8 位表示的 ± 112 ，而不是 ± 0.5 的抽象值。

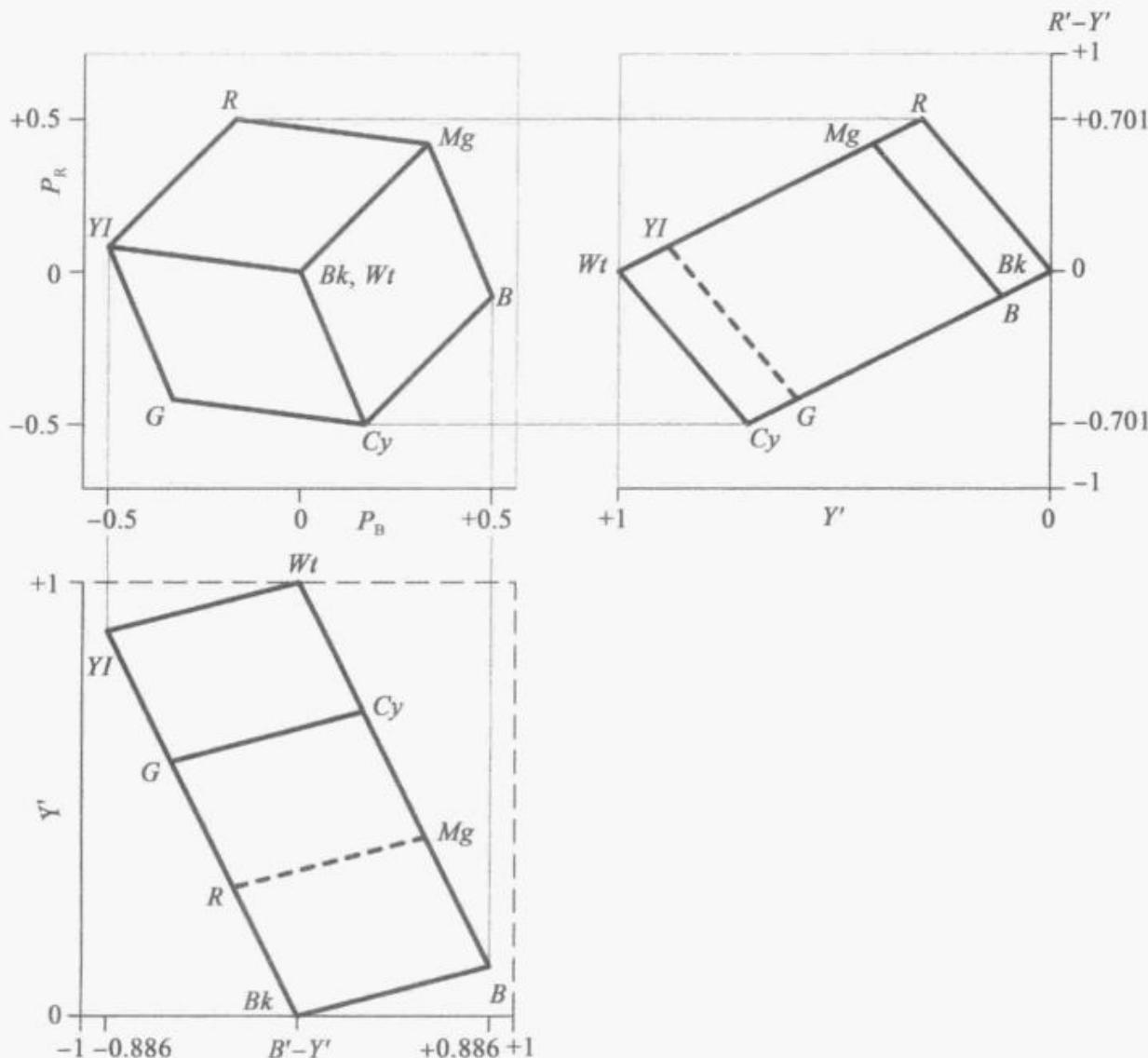


图 28-3 $Y'、B'-Y'、R'-Y'$ 正交视图

28.3 传统的亮度、色差编码

在前文介绍了恒定照度。真正的恒定照度编码仍然有实现的可能性，但是目前所有的视频系统都采用非恒定照度编码，下面对它进行介绍。

传统的亮度、色差编码器如图 28-4 所示。首先，非线性传递函数作用于每个红、绿、蓝（三刺激值）线性分量。然后，经伽马校正后的 R' 、 G' 、 B' 分量的加权和形成亮度。通过减法形成了 $B'-Y'$ 和 $R'-Y'$ 色差分量；图 28-4 显示了缩放到模拟 P_B 和 P_R 分量。最后，对色差分量进行低通滤波。

图 28-4 显示出在亮度通道中有延时元素。亮度的延时与色度经过色度带限滤波器的延时相同。

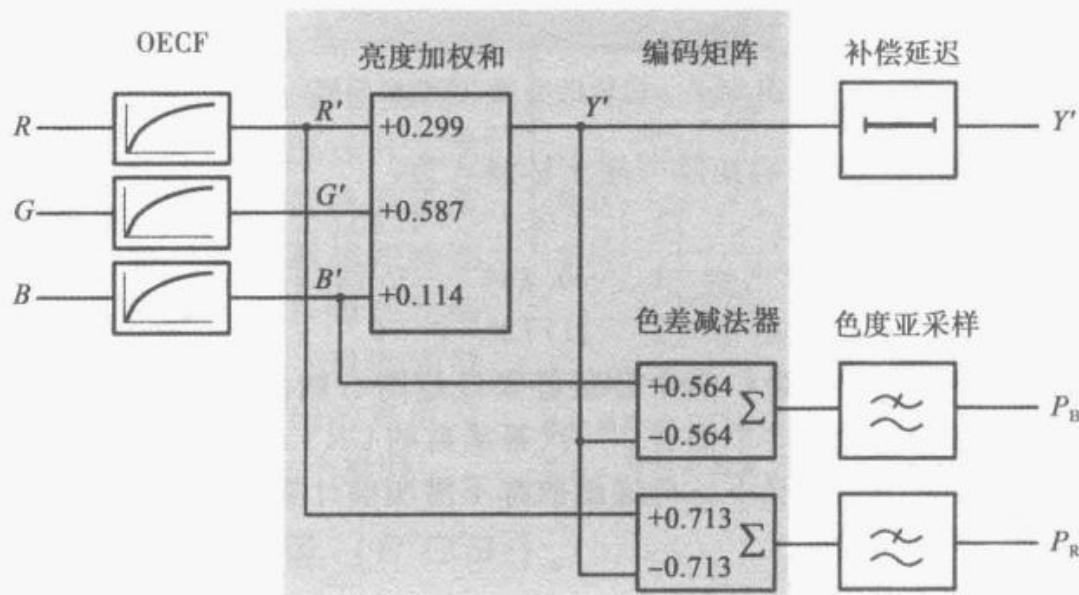


图 28-4 传统的亮度/色差编码器

图 28-4 所示的高亮矩形部分，把形成亮度的加权加法器和一对色差减法器组合在一起，这种组合等效于用式 (28-1) 中的 3×3 矩阵 \mathbf{P} 进行矩阵乘法。在式 (28-1)、图 28-4 以及本章后续插图中采用的数值都反映了用于标清的 BT. 601 亮度系数。可惜，用于高清的系数是不同的，这会在第 30 章介绍。

BT. 601 标准的 $Y'P_BP_R$ 编码矩阵（用于标清）为：

$$\mathbf{P} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.169 & -0.331 & 0.5 \\ 0.5 & -0.419 & -0.081 \end{bmatrix} \quad (28-1)$$

对于该矩阵的推导和更精确的表达，见 29.1 节。

图 28-5 描绘了传统的亮度/色差解码器。在数字解码器中，色差（色度）分量经过水平（在一些应用中，还有空间上）内插；在模拟解码器中，无需电路来完成这个功能。先将亮度与色差分量相加来重建非线性蓝色和红色分量，然后形成亮度、蓝色、红色分量的加权和，重建非线性绿色分量。

在以前的模拟标清电视接收机中，色度插值是隐含的，无需电路实现，在图中色度插值部分用虚线框表示。在以前的模拟标清电视接收机中，EOCF 是 CRT 固有特性，也不需要任何器件。

在图 28-5 所示的高亮矩形中包含的蓝、红色差加法器和还原绿色的权重加法器合起来可看作与式 (28-2) 中的 3×3 矩阵 \mathbf{P}^{-1} 相乘。这些值是用于标清的，而高清的矩阵是不同的。

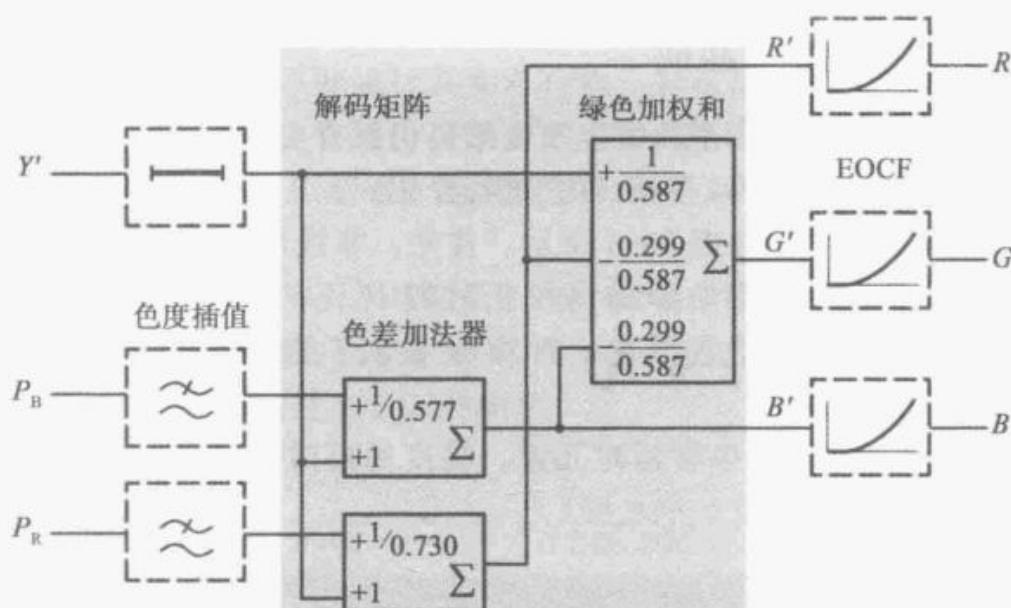


图 28-5 传统的亮度/色差解码器

BT. 601 标准的 $Y'P_BP_R$ 解码矩阵（用于标清）为：

$$\mathbf{P}^{-1} = \begin{bmatrix} 1 & 0 & 1.402 \\ 1 & -0.344 & -0.714 \\ 1 & 1.772 & 0 \end{bmatrix} \quad (28-2)$$

为产生线性光三刺激值分量，三个分量都要经过图右侧虚线框中逆传递函数的变换。通常，与解码器一同使用的是一个固有 2.4 次幂函数的 CRT，或是其他某些集成了 2.4 次幂函数的显示器，在这两种情况下，传递函数都不需明确计算。

28.4 照度和亮度标记

24.4 介绍了同光强度成正比的相对（线性光）照度是怎样以适当的 RGB 加权和算出的。

在视频中，并不计算色彩学的照度，而是用非线性（经伽马校正的） $R'G'B'$ 加权和来计算亮度这个非线性量值。权重值，或亮度系数，与照度系数有关。BT. 601 标准规定的亮度系数已普遍用于标清，但是高清标准引入了新的不一样的权重。本书作者认为，高清的亮度系数没有必要，也不应改变：这增加了演播室和消费类设备中上转换和下转换的复杂度，而在性能或质量上没有改善。

从前的电视标准文件用撇号 ('')，常与表示电压的字母 E 一起来表示包含了伽马校正的分量。例如，以前 E'_R 表示经伽马校正后的红色通道。如今，在视频中进行伽马校正被认为是理所当然，以至于 E 和撇号经常被省略。这给那些试图将视频技术应用于其他领域的人带来了许多困惑。

由于存在多个标准的原色度集，新的系数的引入，以及照度和亮度一直存在混淆，需要一种标记法来区分这许多可能的组合。因为没有标准的标记法，作者只好自己发明了一种，如图 28-6 所示。基本符号是 Y、R、G 或者 B。下标表示规定了原色和白色色度的标准。无撇号的字母表示线性三刺激值分量（R、G 或 B），或相对照度（Y）。撇号（'）表示为非线性（经



图 28-6 照度和亮度标记法

伽马校正的) 分量 (R' 、 G' 和 B')，或者亮度 (Y')。

归一缩放对于本章的分量够用了，后面几章将介绍其他的缩放系数。

对于照度或亮度，前置上标表示规定了所使用权重的标准。以前隐含使用 BT. 601 的权重，但是最近的高清标准如 BT. 709 和 SMPTE ST274 要求使用不同的权重。最后，前置下标表示对信号的总体缩放。如果省略该下标，隐含表示总体缩放到归一值，否则，在数字系统中用如 219、255 或者 874 这样的整数指定黑电平到白电平的幅度，或者用如 661、700 或者 714 这样的数字指定模拟信号幅度的毫伏值。

$Y' C_B C_R$ (或 $Y' P_B P_R$) 的排版是个挑战！图 28-7 描绘了要点。根据图 28-6 所示的习惯，给 Y' 增加了前置上标与下标和一个后置下标。缺少这些元素将无法确定想表示的颜色。只在 Y 上加了撇号。有些作者也给 C_B 和 C_R 加了撇号，但本书作者认为，这种做法既过分又学究气。实际上，已部署的图像编码系统要么是所有三种彩色分量都按感知特性编码，要么是（很少见）都按线性光编码。因为没有“混合”系统（线性光照度与非线性色差，或感知编码的亮度与线性光色差），所以无需用三个撇号。



图 28-7 $Y' C_B C_R$ 的排版

撇号应该一直有，用于区别亮度与色彩学中的照度。 C 表示数字信号， P 表示模拟信号。下标 B 和 R 是标签不是变量，所以应当是正体，而不是斜体。 B 在 R 前面

28.5 非线性红、绿、蓝 ($R'G'B'$)

视频产生于对线性光（三刺激值）RGB 三原色分量的近似，通常用范围从 0（黑色）到 +1（白色）的抽象术语来表示。为了从 RGB 三元组有意义地确定一种颜色，三原色和参考白光的色度特性，例如它们的 CIE $[x, y]$ 色度坐标，必须是已知的。《Colour Sciense for Video》讨论了 RGB 分量的色度属性。在缺少任何具体信息的情况下，使用 BT. 709 的三原色和 CIE D₆₅ 白色点。

第 27 章描述了如何对亮度信息进行非线性编码，以用有限的位数获得良好的感知性能。在一个彩色系统中，该章描述的非线性传递函数分别应用于 RGB 三刺激值的每个分量：由 RGB 三刺激值（线性光）取值的集合计算出三个伽马校正后的原色信号，每个信号近似正比于相应场景三刺激值的平方根。

BT. 709 对明显应施加在摄像机中的 EOCF 进行了标准化。对于大了几个百分点的三刺激值，其编码为：

$$\begin{aligned} R'_{709} &= 1.099R^{0.45} - 0.099 \\ G'_{709} &= 1.099G^{0.45} - 0.099 \\ B'_{709} &= 1.099B^{0.45} - 0.099 \end{aligned} \quad (28-3)$$

然而，“BT. 709”编码的重要之处在于编码的图像数据会在标准显示装置上产生预期的画面外观。BT. 1886 标准化了用于高清的参考 EOCF。编码应通过逆 EOCF 来指定，即

$$\begin{aligned} R'_{709} &= R^{1/2.4} \\ G'_{709} &= G^{1/2.4} \\ B'_{709} &= B^{1/2.4} \end{aligned} \quad (28-4)$$

sRGB 标准编码是类似的，但幂指数使用 1/2.2 而不是 1/2.4。

28.6 BT. 601 亮度

下面的亮度公式是 BT. 601 针对标清的标准，也适用于 JPEG/JFIF（在计算技术领域）及 Exif 图像（在数字摄影中）：

$$^{601}Y' = 0.299R' + 0.587G' + 0.114B' \quad (28-5)$$

正如刚才提到，最初用于视频信号的 E 和撇号随着时间的推移被省略了，这导致了 Y 符号在色彩学和电视技术之间的歧义。

亮度方程的系数基于人类视觉对编码中标准化的每个 RGB 原色的灵敏度。蓝色系数值低是饱和蓝色具有较低亮度的结果。亮度系数也是白色点的函数，更确切地说，是参考白色色度的函数。

原则上，亮度系数应该从原色和白色的色度导出。式 (28-5) 的 BT. 601 亮度系数和 1953 年 NTSC 根据原色和白色点确定并在用的系数相同。消费类显示器实际使用的原色在 NTSC 标准被采纳几年后有所变化。今天在 480i 标清使用的原色与 SMPTE RP145 规定的近似，576i 标清使用的原色与 EBU Tech. 3213 规定的近似（这些原色集略有不同；两组原色集都很匹配 BT. 709 的原色）。尽管原色有变化，480i 和 576i 标清视频的亮度系数都依然保持 1953 年确定的数值不变。由于原色的变化，标清的亮度系数理论上不再与原色匹配。这种不匹配没多大实际意义。

标清原色和亮度系数的失配没多大实际意义，而标清和高清亮度系数之间的不匹配却有很大的实际意义！

28.7 BT. 709 亮度

在从 BT. 709 原色导出的“理论上正确的”亮度系数的基础上，1990 年达成了 BT. 709 国际协议：

$$^{709}Y' = 0.2126R' + 0.7152G' + 0.0722B' \quad (28-6)$$

28.8 再论色度亚采样

色差编码的目的是允许亚采样。在模拟视频中，色差分量要用模拟低通滤波器减小带宽，水平方向的彩色细节被除去。数字视频中，色度分量通过滤波后丢弃采样点来进行亚采样或抽取。“色度亚采样”中图 12-3 显示了几种数字亚采样方案。在 4：2：2 亚采样中，经滤波后，在编码器端交替的色差样值被丢弃。在 4：2：0 中，垂直色度细节也被丢弃。在解码器端，缺失的样值通过内插来近似。

模拟色度限制带宽处理以及数字亚采样处理都丢失了一些彩色细节。然而，由于视觉的彩色分辨率差，在正常观看距离上，观看者是察觉不到颜色损失的。

一些低端数字视频系统只是在编码器端不经过滤波简单地丢弃色度像素，并在解码器端复制色度像素。丢弃采样点可看作点采样，该操作会有引入混叠的风险。应该使用合适的抽取和内插滤波器，应按照“滤波和采样”中说明的原理来设计滤波器。

28.9 亮度/色差总结

当亮度和色差编码用于图像交换，从编码器输入端到解码器输出端保持红、绿、蓝三色特性的一致是非常重要的。原色的色品在“视频的色彩学”有详细介绍，本章因其涉及亮度编解码也有所提及。作者假定原色的特征在整个系统中都是匹配的。亮度和色差编码所依赖的原色称为交换（或传输）原色。

在实践中，摄像机传感器可能会产生色品与交换原色不匹配的 RGB 分量。为了在这样的摄像机中实现精确的色彩再现，需要插入一个 3×3 矩阵，将三刺激值信号从成像捕获的原色变换到交换原色（这是摄像机内置的“线性矩阵”）。同样，解码器也可能需要驱动一个原色与交换原色不同的显示器；在解码器的输出端，可能需要插入一个 3×3 矩阵，将交换原色变换为图像显示原色。

图 28-8 总结了亮度/色差编码。如果图像数据产生于线性 XYZ 分量，会使用 3×3 变换矩阵 T_1^{-1} 得到具有交换原色的色度及参考白色的线性 RGB。对于标清和高清的 BT. 709 交换原色标准，变换矩阵如式 (26-9) 所示。更典型地，图像数据产生于一些依赖于具体设备的空间，标记为 $R_1G_1B_1$ ， 3×3 “线性矩阵”变换 (T_1^{-1}) 由摄像机设计者决定。参考图 26-3 到图 26-8，以及搭配的文字和标题，可了解到这样的矩阵是如何设计出来的。实际的摄像机不具备 CIE 色彩匹配函数的线性组合的光谱敏感度，所以它们不能很好地用色度来表征。尽管如此，一旦选定了一个变换到交换原色集的线性矩阵，式 (26-10) 就可用于导出等效的传感器原色（“摄取原色”）。

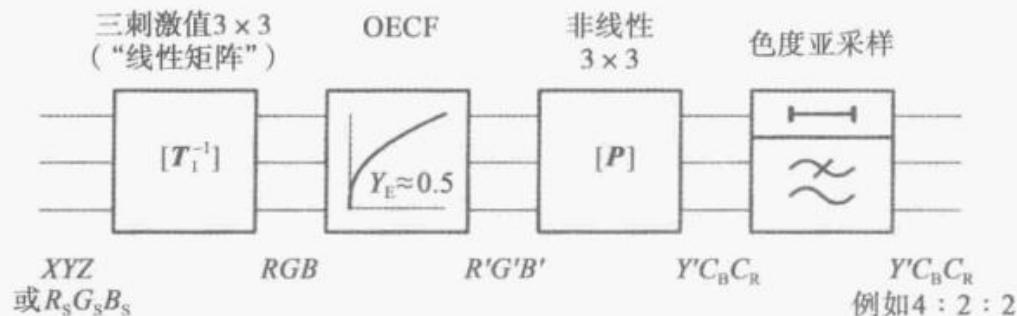


图 28-8 亮度/色差编码器

一旦应用了线性矩阵，每一个分量要经过一个非线性传递函数（伽马校正）产生非线性 $R'G'B'$ 。这些分量经过 3×3 矩阵 (P) 变换得到亮度和色差分量 $Y'C_B C_R$ 或 $Y'P_B P_R$ （此矩阵取决于所使用的亮度系数和色差的比例因子）。然后，如果必需，用色度亚采样滤波器获得亚采样的色差分量；亮度分量有一个补偿延迟。

解码器以相反的顺序进行编码器的逆操作，如图 28-9 所示。在数字解码器中，色度插值滤波器重建丢失的色度样值；在模拟解码器中，不需要显式的操作。 3×3 色差矩阵 (P^{-1}) 重建非线性红、绿、蓝三原色分量。传递函数把原色分量恢复为其线性光三刺激值。最后，三刺激值 3×3 矩阵 (T_2) 从交换标准的原色变换为显示设备中实现的原色。

图 28-8 和图 28-9 给出了用于两个明显不同任务的 3×3 矩阵变换。当有人给你一个 3×3 矩阵，你必须弄清其与光功率是线性还是非线性相关的。

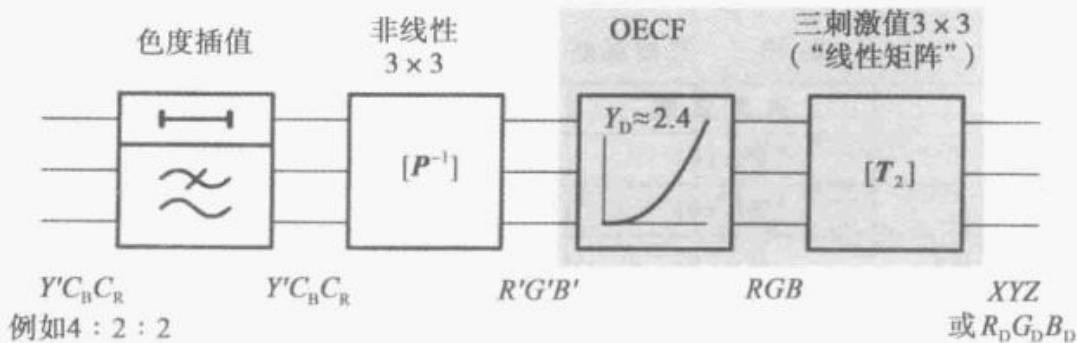


图 28-9 亮度/色差解码器

当解码器与 CRT 显示器联系紧密时，解码器的传递函数用 CRT 固有的非线性电压到照度关系来完成：这个步骤不需要显式的操作。然而，为了利用该传递函数，显示原色必

须与交换原色相同或至少是相似。

解码器（或 CRT）的传递函数与伽玛校正是一起的。一个指数约为 1.2 的端到端的幂函数适合于典型的在演播室照明水平拍摄，并在昏暗环境下以 100nt 亮度观看的电视节目。图 28-8 所示的编码器使用 0.5 次幂函数，图 28-9 所示的解码器使用 2.4 次幂函数。这些产品实现了端到端的幂函数。如果显示设备自身的 EOCF 不同于 CRT 的，解码时应包括一个结合了 2.4 次幂函数和显示设备的逆传递函数的传递函数。

当在一个平均亮度的环境（假设为 20%）中观看一个相当亮的显示屏（假设为 320nt）时，端到端的 1.1 次幂函数是合适的；2.2 次幂 EOCF（类似 sRGB 的 EOCF）是合适的。当在黑暗（0%）环境中观看时，端到端的 1.3 次幂函数是合适的；2.6 次幂 EOCF（类似数码相机的 EOCF）是合适的。

由于最佳的端到端幂函数依赖于观看条件，应该有用于呈现意图的一项用户控制，也许甚至取代亮度和对比度控制，但是并没有。

如果显示原色与交换原色匹配，就不需要解码器的 3×3 三刺激值矩阵。如果显示器的原色与交换原色差不太多，可以非线性域使用一个 3×3 矩阵来补偿原色。不过，如果原色相差很多，就需要在三刺激值域中进行原色之间的变换，见 26.14 节。

28.10 标清和高清亮度的混乱

尽管 $Y'P_B P_R$ 和 $Y'C_B C_R$ 编码的概念在标清和高清中是相同的，BT. 709 标准建立了一套新的高清亮度系数。这套标准与 BT. 601 规定的标清亮度系数明显不同。现在有两种 $Y'C_B C_R$ 编码格式，如图 28-10 所示。这里用⁶⁰¹ $Y'C_B C_R$ 标记标清，用⁷⁰⁹ $Y'C_B C_R$ 标记高清。类似地，有两种 $Y'P_B P_R$ 格式用于模拟系统，⁶⁰¹ $Y'P_B P_R$ 标记标清，⁷⁰⁹ $Y'P_B P_R$ 标记高清。

作者认为采纳不同的编码太不幸了：现在图像编解码要依赖于画面是大（高清）还是小（传统视频、标清）；这种依赖性削弱了在 $Y'C_B C_R 4:2:2$ 和 $4:2:0$ 域中制作节目与分辨率无关的这个非常有用的理念。作者看来，高清应采用 BT. 601 的亮度系数进行标准化。由于有不同的编码，彩色编码参数的混杂使精确的图像交换非常困难。形势很可能随着时间推移而变得更糟而不是更好。

表 28-1 总结了 2011 年左右用于视频解码和显示的原色色品、传递函数，以及亮度系数标准，这些在几个标清和高清标准中或隐含着或很明确。当视频在这些标准之间转换时，为保持期望的颜色，要进行适当的处理。

表 28-1 亮度系数、EOCF 和原色的色品

系 统	亮 度 系 数	EOCF	原 色 色 品
SD 480i	BT. 601	未规定	SMPTE RP 145
SD 576i	BT. 601	未规定	EBU Tech. 3213
HD 720p、1080i、1080p	BT. 709	BT. 1886	BT. 709

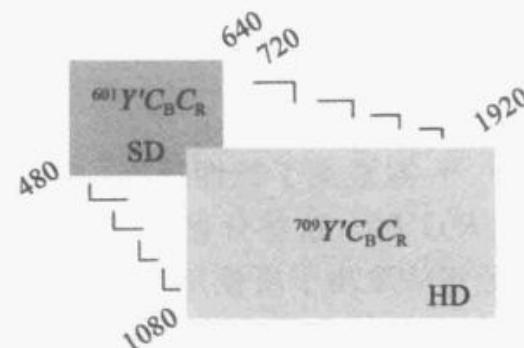


图 28-10 两种亮度/色差格式

注：在所有系统中的编码通过形成能在参考显示器和观看条件下产生预期画面外观的 $R'G'B'$ 值来实现。对于标清，

目前没有有效的关于参考 EOCF 或观看条件的标准。对于高清，BT. 1886 标准化了参考 EOCF，但没有标准化观看条件。

彩条测试信号在 $R'G'B'$ 域进行了标准化，没有参考原色、传递函数和亮度系数。彩条的颜色取决于使用的是哪种原色色度；彩条的亮度和色差电平取决于使用的是哪种亮度

系数。当彩色转换和标准转换正确执行时，彩条测试信号的颜色和电平将改变！

使用术语彩条测试信号（或图案）而不是彩条测试图像是合理的，因为信号是标准化的，而图像不是。

28.11 亮度、色差分量集

在用的有下列色差分量集，都基于 $B'-Y'$ 和 $R'-Y'$ 。

- $Y'P_B P_R$ 编码用于分量模拟视频； P_B 和 P_R 经过缩放具有和亮度相同的标称幅度。 $Y'P_B P_R$ 可能是基于 BT. 601（用于标清）或 BT. 709（用于高清）。在 SD480i29.97 标清中，在用的有三种不同的模拟接口标准：EBU N10 “SMPTE”、索尼、松下。
- $Y'C_B C_R$ 编码用于分量数字视频； C_B 和 C_R 经过缩放具有亮度的 224/219 幅度。在 JPEG/JFIF 中使用一种“满幅度”的变体。 $Y'C_B C_R$ 可能基于 BT. 601 或 BT. 709 亮度系数。

《Composite NTSC and PAL: Legacy Video Systems》的第 5 章详述了另外两个现在已经过时的分量集，其正确使用仅限于复合标清 NTSC 和 PAL。

- $Y'UV$ 分量只适用于复合 NTSC 和 PAL 系统。 $B'-Y'$ 和 $R'-Y'$ 进行缩放来限制复合信号（亮度加上调制的色度）的幅度。 $Y'UV$ 编码一直是基于 BT. 601 亮度系数的。
- $Y'IQ$ 分量历史上从 1953 年到 1970 年左右用于复合 NTSC 系统。UV 分量被旋转 33° 并互换坐标轴，以进行宽带 I 信号传输。这种过时的技术从 1970 年左右起不再使用。 $Y'IQ$ 编码一直是基于现在记载于 BT. 601 文档的亮度系数。

这些让人困惑的比例因子和亮度系数列于表 28-2A（模拟标清）、表 28-2B（数字标清和计算系统）和表 28-2C（模拟和数字高清）中。之后的两章将分别介绍标清和高清的分量彩色编码。

在表 28-2A 第一行显示的 EBU N10 的电平合理但不流行。480i 标清演播室设备的设计者需要实现三类接口“标准”的配置：EBU N10（“SMPTE”）、索尼、松下。

表 28-2A 模拟标清的色差系统制式

系统制式	标记	色差缩放
(1) 分量模拟视频，480i (EIA/CEA-770 和 “SMPTE”) 和 576i EBU N10，以及零建立电平的 480i 松下 M-II (用于日本)	$\frac{601}{700} Y'_{145} P_B P_R$	EBU N10 标准规定了 7:3 的画面-同步比、700mV 亮度幅度、零建立电平。 P_B 和 P_R 分量各自缩放到 $\pm 350\text{mV}$ 的范围，与亮度相同
(2) 分量模拟视频，7.5% 建立电平的 480i 索尼	$\frac{601}{661} Y'_{145} P_B P_R$	索尼事实上的标准规定了 10:4 的画面-同步比、7.5% 建立电平，黑到白的亮度幅度约为 661mV。 P_B 和 P_R 分量各自缩放到 $4/3$ 倍 $\pm 350\text{mV}$ 的范围，即 $\pm 466 \frac{2}{3}\text{mV}$
(3) 分量模拟视频，零建立电平的 480i 索尼 (用于日本)	$\frac{601}{714} Y'_{145} P_B P_R$	索尼事实上的标准规定了 10:4 的画面-同步比，零建立电平，黑到白的亮度幅度约为 714mV。 P_B 和 P_R 分量各自缩放到 $4/3$ 倍的 $\pm 350\text{mV}$ 的范围，即 $\pm 466 \frac{2}{3}\text{mV}$
(4) 分量模拟视频，7.5% 建立电平的 480i 松下 ^①	$\frac{601}{647} Y'_{145} P_B P_R$	松下事实上的标准规定了 7:3 的画面-同步比、7.5% 建立电平，黑到白的亮度幅度约为 647.5mV。 P_B 和 P_R 分量各自缩放到 $37/40$ 倍的 $\pm 350\text{mV}$ ，即 $\pm 323.75\text{mV}$

(续)

系统制式	标记	色差缩放
(5) 复合模拟 NTSC、PAL 视频（包括 S-Video）	多种形式，典型的有： $\frac{601}{700}Y'_{EBU}UV$ 、 $\frac{601}{714}Y'_{145}IQ$ 、 $\frac{601}{714}Y'_{145}UV$	U 和 V 经过缩放以满足一个联合约束：缩放使得峰值复合视频（亮度加色调制色度）限制在消隐到白色幅度的 $4/3$ 。在模拟编码后无法区分坐标轴的旋转和互换（例如 I 和 Q）。没有标准的分量接口

①消费类设备（如 DVD 播放机）的分量模拟接口是按 EIA/CEA-770.2 正确的 $Y'P_BP_R$ 缩放。一些消费类设备在设计和部署时用了不正确的 $Y'P_BP_R$ 缩放。某些用户设备的后面板接口标记着 Y、B-Y、R-Y 或 YUV，这些名称明显是错误的。

在表 28-2B 中，建议使用第一行所示的缩放。关于过时的标清系统的详细信息，见《Composite NTSC and PAL: Legacy Video Systems》第 14 章中的表格“用于模拟分量标清和数字 $4f_{SC}$ 标清的色差系统”（表 28-2B 中的行编号是不连续的，以对应上面提到的表格）。

表 28-2B 用于数字标清和计算技术的色差系统

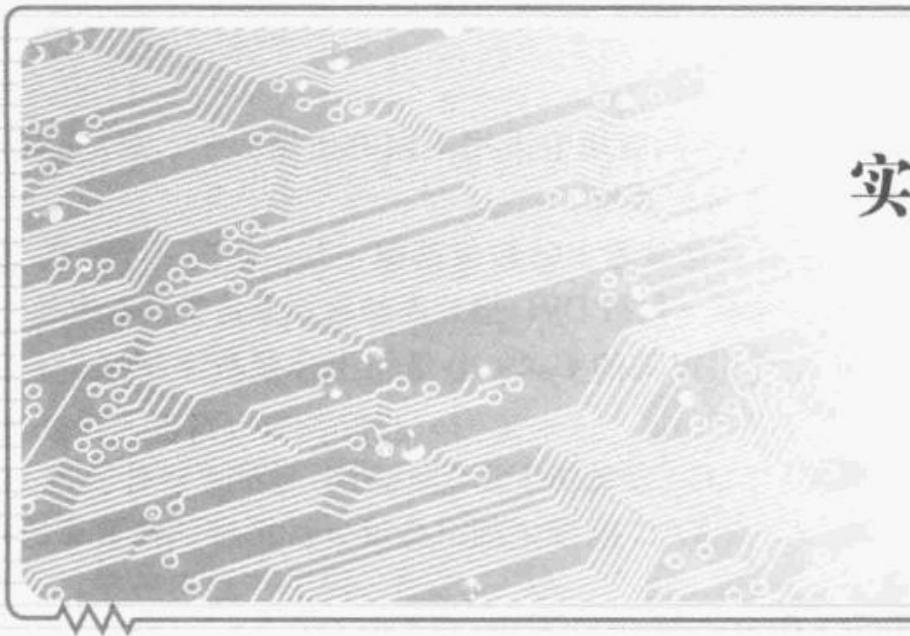
系统	标记	色差缩放
(6) 分量数字视频：4:2:0, 4:1:1, BT.601 4:2:2（包括 M-JPEG、MPEG、DVD、DVC）	$\frac{601}{219}Y'_{145}C_B C_R$	BT.601 规定了亮度范围为 0~219，在接口上有 +16 的偏置。 C_B 和 C_R 分量各自缩放到 ±112 的范围，幅度是亮度的 $224/219$ 倍，在接口上有 +128 的偏置。代码 0 和 255 被禁用
(7) 分量数字静止帧 JPEG（包括 JFIF 1.02），典型的桌面印刷和网页。传递函数不相同。	$\frac{601}{255}Y'_{709}C_B C_R$	没有综合性的标准。亮度参考范围典型地为 0~255。 C_B 和 C_R 典型地缩放到 ±128 的“全摆幅”，幅度是亮度的 $256/255$ 倍。 C_B 和 C_R 的代码 +128 被钳位；无法表示完全饱和的蓝色和红色

表 28-2C 用于高清的色差系统

系统	标记	色差缩放
(11) 分量模拟高清：4:2:0, 4:1:1, BT.601 4:2:2（包括 M-JPEG、MPEG、DVD、DVC）	$\frac{709}{700}Y'_{709}P_BP_R$	7:3 的画面 - 同步比、700mV 亮度幅度、零建立电平。 P_B 和 P_R 分量各自缩放到 ±350mV 的范围，与亮度相同
(12) 分量数字高清（BT.709/BT.1886）	$\frac{709}{219}Y'_{709}C_B C_R$	BT.601 规定了亮度范围为 0~219，在接口上有 +16 的偏置。 C_B 和 C_R 分量各自缩放到 ±112 的范围，幅度是亮度的 $224/219$ 倍，在接口上有 +128 的偏置。代码 0 和 255 被禁用
(13) 分量数字高清（xvYCC）	$\frac{xvYCC}{219}Y'_{709}C_B C_R$	xvYCC 的 $Y'C_B C_R$ 与 BT.709 的相同，除了有些位于 $R'G'B'$ 单位立方体之外的码字是表示宽色域颜色

○ 第三部分

实际问题



第 29 章

标清分量视频彩色编码

基本色差分量 $B'-Y'$ 和 $R'-Y'$ 在不同应用中采用不同的比例因子。第 28 章介绍了亮度和色差编码。本章将详细介绍以下编码系统。

- $B'-Y'$ 、 $R'-Y'$ 分量形成所有其他分量集的数值基础；其他情况下不直接使用。
- $P_B P_R$ 分量用于分量模拟视频（包括 DVD 播放器和机顶盒这类设备中的模拟接口）。
- 在 BT. 601 和 BT. 709 标准中定义的 $C_B C_R$ 分量用于分量数字视频，包括演播室视频、DV、MPEG 和 H. 264。
- “全摆幅”（或“全范围”） $C_B C_R$ 分量用于 JPEG/JFIF 标准。
- UV 分量用于复合 NTSC 或 PAL 标准，如《Composite NTSC and PAL: Legacy Video Systems》第 5 章所述。
- IQ 分量历史上用于复合 NTSC 标准，直到 1970 年左右，如《Composite NTSC and PAL: Legacy Video Systems》第 4 章所述。

$Y'UV$ 和 $Y'IQ$ 是形成复合 NTSC、PAL 和 S-视频的中间量。 $Y'UV$ 和 $Y'IQ$ 都没有标准分量接口，而且当各分量保持分开时二者均不合适。不幸的是， $Y'UV$ 这个术语的使用比较随意，有些人使用 $Y'UV$ 表示 $B'-Y'$ 和 $R'-Y'$ 的任何缩放。

视频使用符号 U 和 V 来表示某种色差分量。CIE 定义了 $[u, v]$ 、 $[u', v']$ 和 $[u^*, v^*]$ 这几对。所有这些都表示色彩或色度信息，但是它们在数值和功能上均不相同。视频 $[U, V]$ 分量既不是直接基于，也并没有被任何 CIE 彩色空间取代。

本章所描述的编码系统可适用于各种 RGB 原色集，如 EBU 3213、SMPTE RP 145（或甚至可用于 BT. 709）。BT. 601 不指定原色色度：SMPTE RP 145 原色在 480i 标清中是隐含的，EBU 3213 原色在 576i 标清中是隐含的。然而，实际上所有的现代消费类接收机解释无论是标清还是高清内容都会根据 BT. 709 原色来确定。在作者写作本书时，在北美创建的节目内容是按 SMPTE 原色加工的（与 ITU-R、SMPTE 和 ATSC 标准的精神和文字相反），在欧洲创建的内容加工为 EBU 原色。不过，所有这些内容消费者领域都用 BT. 709 原色显示。期待有一天内容创作者真正使用 BT. 709 彩色空间进行加工。

$[Y', B'-Y', R'-Y']$ 、 $Y'P_B P_R$ 和 $Y'C_B C_R$ 的公式可以是基于 BT. 601 标清亮度系数或 BT. 709 高清系数的。本章的公式和插图是基于 BT. 601 的系数的。不幸的是，标准化了的高清亮度系数不同于 BT. 601 的系数。关于高清亮度系数，见 28.7 节；关于高清色差分量的细节，见第 30 章。

色度分量的正确排列为先 $B'-Y'$ 后 $R'-Y'$ ，或先 P_B 后 P_R ，或先 C_B 后 C_R 。蓝色分量与 U 有关，红色分量与 V 有关；U 和 V 按字母排序。 $C_B C_R$ 和 $P_B P_R$ 的下标经常用小写。作者认为这不利于可读性，所以将其大写。 C_B 的 B 作为一个标签而不是变量，所以用罗马字体表示（即竖直而非斜体）。非常注意细节的作者有时给 $C_B C_R$ 和 $P_B P_R$ 加撇号表示它们来自非线性光，但标准的或已部署的图像编码系统都不采用线性光色差，也没有感觉方面的理由要加撇号，所以这里省略了撇号。

29.1 标清的 $B'-Y'$ 、 $R'-Y'$ 分量

为了从 $R'G'B'$ 得到 $[Y', B'-Y', R'-Y']$ ，对于 BT. 601 亮度，使用下面的矩阵公式：

$$\begin{bmatrix} {}^{601}Y' \\ B' - {}^{601}Y' \\ R' - {}^{601}Y' \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.299 & -0.587 & 0.886 \\ 0.701 & -0.587 & -0.114 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (29-1)$$

图 29-1 显示了 $[B'-Y', R'-Y']$ 色差平面图。

如 28.6 所述，BT. 601 的亮度系数用于标清。用这些系数时， $B'-Y'$ 分量在纯蓝色 ($R'=0, G=0, B=1, Y'=0.114; B'-Y'=0.886$) 时达到最大正值，在纯黄色 ($B'-Y'=-0.886$) 时达到最大负值。类似地， $R'-Y'$ 极值在纯红色和青色时取值 ± 0.701 。这些值对数字和模拟系统都不便于使用。 $P_B P_R$ 、 $C_B C_R$ 和 UV 色差分量都涉及不同版本的 $[Y', B'-Y', R'-Y']$ ，经过缩放使各分量的极值便于使用。

ITU-R Rec. BT. 601-5，“标准 4:3 和宽屏幕 16:9 宽高比数字电视的演播室编码参数”。

29.2 标清的 $P_B P_R$ 分量

P_B 和 P_R 表示色差分量，其标称幅度与伴随的亮度分量幅度相同。对于 BT. 601 亮度，方程如下：

$$\begin{aligned} P_B &= \frac{0.5}{1 - 0.114} (B' - {}^{601}Y') = \frac{1}{1.772} (B' - {}^{601}Y') \approx 0.564 (B' - {}^{601}Y') \\ P_R &= \frac{0.5}{1 - 0.299} (R' - {}^{601}Y') = \frac{1}{1.402} (R' - {}^{601}Y') \approx 0.713 (R' - {}^{601}Y') \end{aligned} \quad (29-2)$$

选中的这些比例因子将各色差分量的幅度限制在相对于归一化的亮度幅度为 -0.5 到 $+0.5$ 的范围内：在上面第一个表达式中 0.114 是蓝色分量的亮度系数，第二个表达式中 0.299 是红色分量的系数。图 29-2 显示了 $[P_B, P_R]$ 平面图。

用矩阵形式表示时，式 (29-1) 中的 $B'-Y'$ 和 $R'-Y'$ 行分别缩放 $0.5/0.886$ 和 $0.5/0.701$ 倍。从参考黑色为 0 且参考白色为归一值的 $R'G'B'$ 进行编码：

$$\begin{bmatrix} {}^{601}Y' \\ P_B \\ P_R \end{bmatrix} = \begin{bmatrix} 0.299\ 000 & 0.587\ 000 & 0.114\ 000 \\ -0.168\ 736 & -0.331\ 264 & 0.500\ 000 \\ 0.500\ 000 & -0.418\ 688 & -0.081\ 312 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (29-3)$$

式 (29-3) 的第一行由亮度系数组成，其和为归一值。第二和第三行各自加起来都为 0，这对于色差分量是必需的。两个为 0.5 的项反映了 P_B 和 P_R 在蓝色和红色原色 $[0, 0, 1]$ 和 $[1, 0, 0]$ 时的参考幅度。参考幅度为 ± 0.5 ；峰值幅度可能稍大一些，以适应模拟下冲和上冲。没有标准说明应提供多大的模拟下余量和上余量。

解码的逆矩阵为：

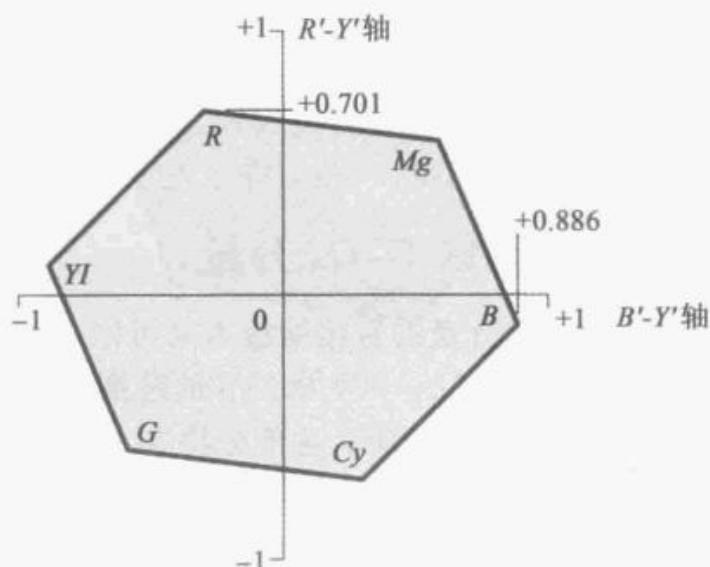


图 29-1 标清的 $B'-Y'$ 、 $R'-Y'$ 分量

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1.000\ 000 & 0.000\ 000 & 1.402\ 000 \\ 1.000\ 000 & -0.344\ 136 & -0.714\ 136 \\ 1.000\ 000 & 1.772\ 000 & 0.000\ 000 \end{bmatrix} \cdot \begin{bmatrix} {}^{601}Y' \\ P_B \\ P_R \end{bmatrix} \quad (29-4)$$

$Y'P_BP_R$ 应用于480i和576i分量模拟视频设备，如索尼和松下的产品，其中 P_B 和 P_R 用差不多亮度的一半带宽传送。不幸的是，在用的有三种不同的模拟接口电平标准： $Y'P_BP_R$ 的电气接口有歧义。

P_B 和 P_R 按照第29章描述的顺序书写。 P 代表并行，源于SMPTE内部的一次将分量模拟视频的并行电气接口标准化的失败尝试。在现在要讲到的 C_B C_R 中， C 代表色度。 C_B C_R 符号先于 P_B P_R 。

29.3 标清的 C_B C_R 分量

$Y'P_BP_R$ 分量的直接缩放本来可以适合数字接口。

将亮度缩放到 $[0, \dots, 255]$ 的范围本来是可行的；这种亮度的“全范围”缩放用于计算领域的JPEG/JFIF，这部分将在29.6节介绍。然而，在演播室应用中，需提供信号处理的下余量和上余量，以容纳来自模拟和数字滤波器的振铃，以及容纳来自被错误调节的模拟设备的信号。

对于8位接口，亮度本来可以缩放到224的幅度； $B'-Y'$ 和 $R'-Y'$ 本来可以缩放到±112。这样就会为每个分量留下了32个编码的下余量和上余量。尽管这种方法是明智的，但在1984年通过BT.601时并没有被采纳。相反，而且按作者的观点是不幸的，不同的用于亮度和色度的幅度被标准化。8位亮度幅度标准化为219，色度幅度标准化为224。每个色差分量的幅度是亮度幅度的 $224/219$ 倍。由于视频分量幅值一般参考亮度幅度，这种情况反过来说更清楚：在 $Y'P_BP_R$ 中，每个色差分量有 $224/219$ 倍亮度分量幅度。 C_B C_R 的亮度和色度的幅度名义上是相同，这一点与 P_B P_R 不一样：概念上说， $Y'P_BP_R$ 和 $Y'C_B$ C_R 仅在缩放比例上有区别。

历史上， $Y'P_BP_R$ 缩放用于模拟接口， $Y'C_B$ C_R 用于数字接口。当今有如此多不同的缩放因子和幅度同时在模拟域和数字域使用，以至于双命名法的麻烦多于帮助。

$Y'P_BP_R$ 和 $Y'C_B$ C_R 的缩放比例不一致已经够不幸了，而用于JPEG/JFIF的比例类似 $Y'P_BP_R$ 但不相同的“全摆幅”（或“全范围”的 $Y'C_B$ C_R 让情况更糟。EBU在技术标准N10-1998中把应该标记为 P_B P_R 的信号称为 C_B C_R 模拟色差分量，也造成了混乱。

为了提供下余量来容纳略有负值的亮度信号，在 $Y'C_B$ C_R 接口上给亮度增加了一个偏移量。在8位接口加上+16的偏移量，这将黑色用代码16表示，白色用代码235表示。在8位接口上，代码0~255用于同步；这些代码对视频数据禁用。代码1~15解释为相对于归一化的亮度幅度从 $-15/219$ 到 $-1/219$ 的信号电平；代码236到254解释为相对于归一化幅度从 $220/219$ 到 $238/219$ 的信号电平。不幸的是，亮度的下余量和上余量是不对称的。

C_B C_R 色差分量以偏移二进制形式传输：加入一个+128的偏移量。在演播室 $Y'P_BP_R$ 中，色度参考电平的代码是16和240，并且码0和255禁止为色度数据。

BT.601支持10位分量；10位演播室视频设备现在很常见。在10位接口上，8位接口电平和禁用码不变；多出的位作为最低有效位（LSB）提供更高的精度。禁用码遵守8位接口习惯：在10位接口上，8位最高有效位全为0或全为1的代码对视频数据是禁用的。

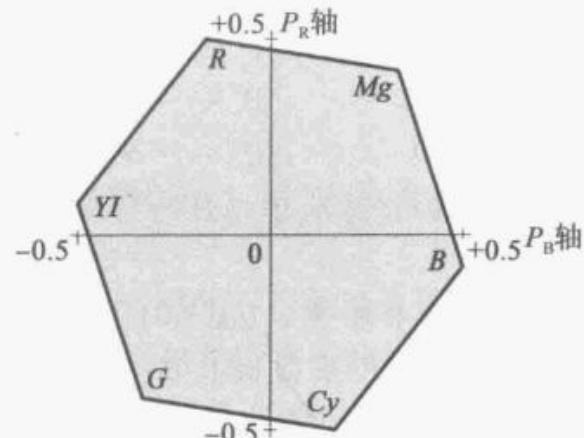


图29-2 标清的 P_B P_R 分量

10位接口的参考白色和参考黑色的代码是在8位数据的最低有效位之后补零。从8位“拓宽”到更高精度应当通过左移实现，而不是乘879/219。“缩窄”到8位应通过舍入后右移来实现。

对于信号处理算术运算，如增益调节，黑色的 Y' 、 C_B 和 C_R 必须为零，一定要移除接口偏移量。对于8位亮度算术运算，可以很方便地把参考黑色用代码0表示，把参考白色用代码219表示。色差信号在2的补码形式下最便于处理，通过缩放使参考色差信号（纯黄色、青色、红色和蓝色）因子为±112。图29-3显示了用这种方式缩放得到的 $C_B C_R$ 色差平面，没有偏移量。

就作者而言，可认为偏移是一个接口。尽管如此，大部分对 $Y' C_B C_R$ 的描述，包括SMPTE和ITU标准，都认为 $Y' C_B C_R$ 符号包括偏移。在下面的公式，偏移量用黑体标记。如果是为了计算用于有符号数信号处理的抽象数学量，应省略这些偏移量。如果关心的是接口的无符号值，应当包括偏移量。

这些公式从范围为 $[0, \dots, 1]$ 的 $[Y', B'-Y', R'-Y']$ 分量形成BT.601 $Y' P_B P_R$ 分量：

$$\frac{601}{219} Y' = \mathbf{16} + (219 \times {}^{601} Y')$$

$$C_B = \mathbf{128} + \frac{112}{0.886} \times (B' - {}^{601} Y') \quad (29-5)$$

$$C_R = \mathbf{128} + \frac{112}{0.701} \times (R' - {}^{601} Y')$$

这个公式以及之后的公式中使用的数值是以BT.601的亮度系数为基础的。不幸的是，用于高清的系数是不同的，见28.6节。

为了把式(29-5)扩展到10位，给每个 Y' 、 C_B 和 C_R 附加2个二进制权重为1/2和1/4的低阶位。为了把 $Y' P_B P_R$ 扩展到超过10位，在位序列后加上权重为1/8、1/16等等的最低有效位。如果愿意用整数表示这些量，不带小数位，可以用 2^{K-8} 乘式(29-5)(以及之后所有的公式)，其中 $8 \leq k$ 表示位深度。

为了从范围为0到1的 $R' G' B'$ 得到8位BT.601 $Y' P_B P_R$ ，用因子219、224和224对式(29-3)中矩阵的各行进行缩放，分别对应 Y' 、 C_B 和 C_R 的幅度：

$$\begin{bmatrix} \frac{601}{219} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} \mathbf{16} \\ \mathbf{128} \\ \mathbf{128} \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (29-6)$$

把矩阵最上面一行相加得到219，是亮度幅度。下面两行相加为零。两个为112的项表示在蓝和红原色上的 C_B 和 C_R 正极值。

要从8位BT.601 $Y' P_B P_R$ 恢复范围是 $[0, \dots, +1]$ 的 $R' G' B'$ ，可反转式(29-6)，得到：

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 0.00456621 & 0 & 0.00625893 \\ 0.00456621 & -0.00153396 & -0.00318811 \\ 0.00456621 & 0.00791071 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{601}{219} Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} \mathbf{16} \\ \mathbf{128} \\ \mathbf{128} \end{bmatrix} \quad (29-7)$$

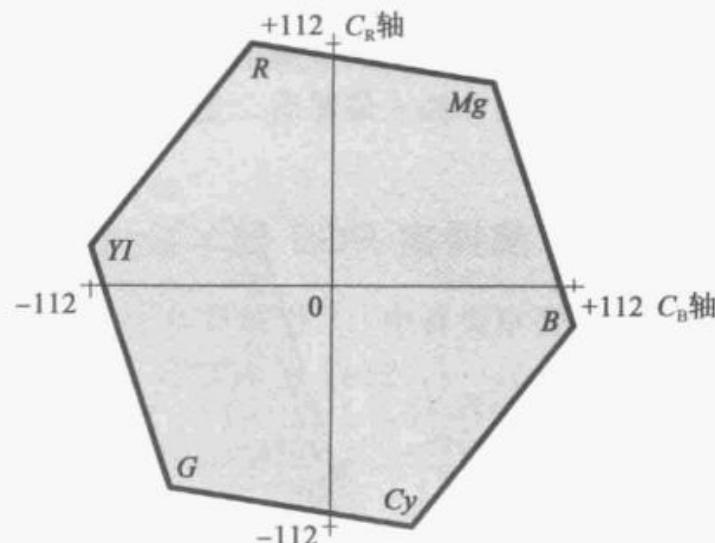


图29-3 标清的 $C_B C_R$ 分量

这个矩阵包含大于 256 的项目，对应的乘法器需要超过 8 位的容量。

你可以通过对矩阵第一行系数求和来确定编码矩阵设计时要产生的幅度——经常是 1、219、255 或 256。在式 (29-8) 中和为 256。如果和不是预计的值，可怀疑矩阵中有错误。

当对矩阵系数进行舍入时，注意保持设定的行总和，在这种情况下为 [1, 0, 0]。需注意防止由于舍入错误或其他条件产生的溢出：使用饱和算法。

在接口上，加入偏移后，把所有三个分量钳位到 1~254 的范围，避免出现禁用代码 0 和 255。

29.4 从演播室 RGB 到 $Y'C_B C_R$

在演播室设备中，8 位 $R'G'B'$ 分量通常和 $Y'C_B C_R$ 的亮度分量有同样 219 的幅度。为从范围为 $[0, \dots, 219]$ 的 $R'G'B'$ 编码 8 位 BT. 601，用 256/129 缩放式 (29-6) 的编码矩阵：

$$\begin{bmatrix} {}^{601}_{219} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 76.544 & 150.272 & 29.184 \\ -43.366 & -85.136 & 128.502 \\ 128.502 & -107.604 & -20.898 \end{bmatrix} \cdot \begin{bmatrix} {}^{219} R' \\ {}^{219} G' \\ {}^{219} B' \end{bmatrix} \quad (29-8)$$

对于二进制算术的实现，乘以 1/256 可以通过移位来完成。从 8 位 BT. 601 $Y'C_B C_R$ 解码到范围是 $[0, \dots, 219]$ 的 $R'G'B'$ ，反转式 (29-8)，得到：

$$\begin{bmatrix} {}^{219} R' \\ {}^{219} G' \\ {}^{219} B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 256 & 0 & 350.901 \\ 256 & -86.132 & -178.738 \\ 256 & 443.506 & 0 \end{bmatrix} \cdot \begin{bmatrix} {}^{601}_{219} Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (29-9)$$

这个矩阵中为 256 的项表示对应的分量可以简单地加上，无需进行乘法运算。这些变换假定 $R'G'B'$ 分量包含了伽马校正，比如 BT. 709 规定的 $R'G'B'$ 。

29.5 从计算机 RGB 到 $Y'C_B C_R$

在计算领域，习惯使用没有上余量和下余量的 8 位 $R'G'B'$ 分量：黑色是代码 0，白色是代码 255。从这个范围的 $R'G'B'$ 编码 8 位 BT. 601 $Y'C_B C_R'$ ，用 256/255 缩放式 (29-6) 的矩阵：

$$\begin{bmatrix} {}^{601}_{219} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 65.738 & 129.057 & 25.064 \\ -37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \cdot \begin{bmatrix} {}^{255} R' \\ {}^{255} G' \\ {}^{255} B' \end{bmatrix} \quad (29-10)$$

从 8 位的 BT. 601 $Y'C_B C_R$ 解码到范围为 $[0, \dots, 255]$ 的 $R'G'B'$ ，使用的变换式为：

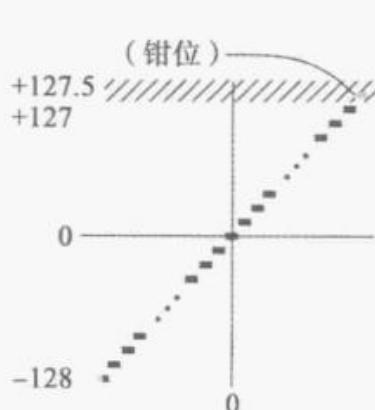
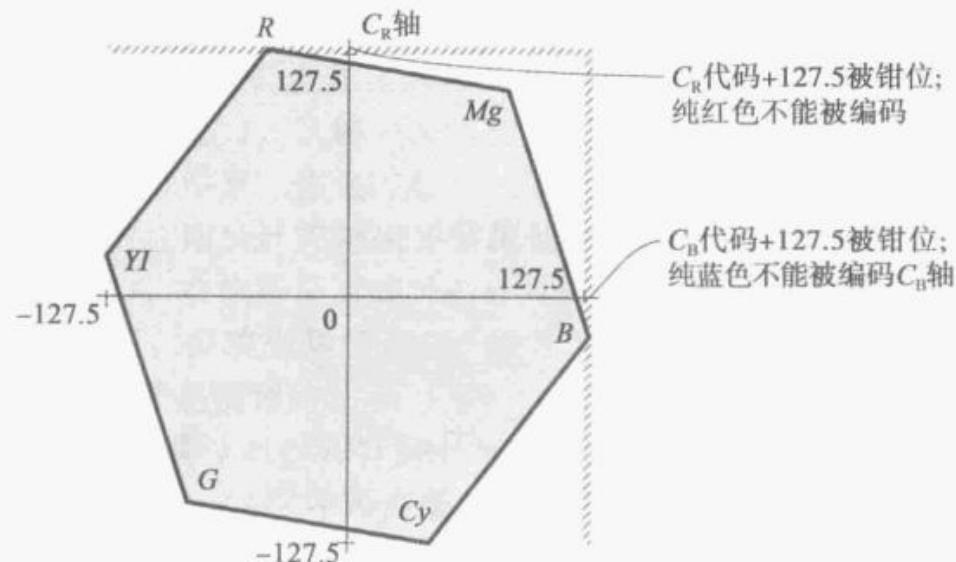
$$\begin{bmatrix} {}^{255} R' \\ {}^{255} G' \\ {}^{255} B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 298.082 & 0 & 408.583 \\ 298.082 & -100.291 & -208.120 \\ 298.082 & 516.411 & 0 \end{bmatrix} \cdot \begin{bmatrix} {}^{601}_{219} Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (29-11)$$

BT. 601 $Y'C_B C_R$ 使用编码范围的极值来处理信号上冲和下冲。当解码到没有上余量和下余量的 $R'G'B'$ 范围时，需要钳位。

29.6 “全摆幅” $Y'C_B C_R$

在计算领域的 JPEG/JFIF 静止帧中使用的 $Y'C_B C_R$ 编码，习惯上采用没有下余量和上余量的“全摆幅”编码。亮度 (Y') 扩展到 255 的幅度，用 8 位表示：黑色用代码 0 表示，白色用代码 255 表示。显然，亮度代码 0 和 255 是不被禁止的！色差分量缩放到

±127.5 的幅度；名义上每个色差分量和亮度具有相同的幅度。然而，加上了+128 的偏移量（而不是所期望的+127.5），显然这样纯灰色就有整数代码值。+128 的偏移量导致纯蓝色和纯红色在偏置后的值为 255.5，这会被钳位。图 29-4 显示了色差量化器的传递函数，强调在偏置前的色度代码+127.5（纯蓝色或纯红色）导致钳位。因此，纯蓝色和纯红色倾向于无法在 JPEG/JFIF 压缩和解压缩之间准确地“往返”。图 29-5 显示了全范围 $C_B C_R$ 色差平面。

图 29-4 “全摆幅” $C_B C_R$ 量化器图 29-5 $C_B C_R$ “全范围” 分量

从范围是 $[0, \dots, 255]$ 的 $R'G'B'$ 编码成 8 位 $Y'C_B C_R$ ，同时亮度范围是 $[0, \dots, 255]$ ， C_B 和 C_R 各自的范围是 ±128，使用变换式为：

$$\begin{bmatrix} \frac{601}{255} Y' \\ C_B \\ C_R \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 76.544 & 150.272 & 29.184 \\ -43.366 & -85.136 & 128.502 \\ 128.502 & -107.604 & -20.898 \end{bmatrix} \cdot \begin{bmatrix} 255 R' \\ 255 G' \\ 255 B' \end{bmatrix} \quad (29-12)$$

从全范围 8 位 $Y'C_B C_R$ 解码到范围是 $[0, \dots, 255]$ 的 $R'G'B'$ ，使用的变换式为：

$$\begin{bmatrix} 255 R' \\ 255 G' \\ 255 B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 256 & 0 & 357.510 \\ 256 & -87.755 & -182.105 \\ 256 & 451.860 & 0 \end{bmatrix} \cdot \begin{bmatrix} \frac{601}{255} Y' \\ C_B \\ C_R \end{bmatrix} \quad (29-13)$$

29.7 $Y'UV$ 、 $Y'IQ$ 的混淆

文中已详细介绍了 $Y'P_B P_R$ 和 $Y'C_B C_R$ 。这些都是基于 $[B'-Y', R'-Y']$ 分量，但是它们分别具有不同的适用于分量模拟接口和分量数字接口的比例因子。

色差对 $[U, V]$ 和 $[I, Q]$ 也基于 $B'-Y'$ 和 $R'-Y'$ ，但有另外的缩放因子。UV 缩放或 IQ 缩放与旋转只在信号用于如 NTSC 或 PAL 的复合编码时适用。

不幸的是，标记 $Y'UV$ ，更糟的是 YUV ，有时被随意地应用到基于 $[B'-Y', R'-Y']$ 的任何形式的色差编码。不要被视频设备误导，这些设备的接头标有 $Y'UV$ 或 Y' 、 $B'-Y'$ 、 $R'-Y'$ ，或不带撇号的这些符号；也不要被描述为使用 $Y'UV$ 编码的 JPEG 误导。实际上，对于模拟接口传输带 $Y'P_B P_R$ 缩放的信号，JPEG 标准本身规定了在本书中表示为 $\frac{601}{255} Y'P_B P_R$ 的信号格式。

当在计算机图形学或图像处理的上下文中碰到术语 $Y'UV$ （或 YUV ）时，通常指的是 BT. 601 $Y'C_B C_R$ ，但要当心！

- 任何据称编码到原始的 1953 NTSC 原色的图像数据都值得怀疑，因为已经有差不

多 40 年都没有任何使用这些原色的设备生产出来。

- 通常不提及底层 $R'G'B'$ 分量的传递函数，而且没有考虑亮度的非线性计算。
当遇到术语 $Y'IQ$ （或 YIQ ），要当心！
- 声称是用 $Y'IQ$ 编码的图像数据值得怀疑，因为没有用于 $Y'IQ$ 分量的模拟或数字接口曾被标准化。
- 自 1970 年以来搭建的 NTSC 编码器和解码器是基于 $Y'UV$ 分量，而不是基于 $Y'IQ$ 的。与许多发布的信息相反， $Y'IQ$ 分量已有约 40 年未用于“NTSC”了。

第30章

高清分量视频彩色编码

第29章详细介绍了各种使用BT.601指定亮度系数的分量彩色编码系统。不幸的是，不是出于好的技术原因，用于高清的BT.709标准化了不同的亮度系数。部署高清要求演播室和消费者家里都具备上变换和下变换能力，这种情况将会持续几十年。由于高清和标清之间需要转换，如果你想成为一个高清专家，你也需要是一个标清专家！

如今用于静止帧、桌面视频和其他应用的计算机成像系统通常使用BT.601参数，独立于图像的像素数（“分辨率独立性”）。在执行高清编辑的计算机系统中，非常希望同一时间轴上所有内容使用相同的彩色编码，但应采用BT.601还是BT.709编码，并没有简单的答案。一般来说，保持BT.601系数是明智的。

在BT.709制定之前，用于1035i30高清的SMPTE240M-1988标准化了基于SMPTE RP145原色的亮度系数。在大约1988年和1997年间部署的设备使用的是240M参数，但SMPTE 240M现在已经过时了。详细信息见本书第1版。

假定读者已经熟悉“亮度和色差”的概念，本章将详细介绍以下分量集：

- $B'-Y'$ 、 $R'-Y'$ 分量， $P_B P_R$ 和 $C_B C_R$ 的基础。
- $P_B P_R$ 分量，用于模拟接口。
- $C_B C_R$ 分量，用于数字接口。

30.1 BT.709高清 $B'-Y'$ 和 $R'-Y'$ 分量

$B'-Y'$ 分量在蓝色($R'=0, G'=0, B'=1$)时达到其最大正值。用BT.709的亮度系数， $B'-Y'$ 的最大值为+0.9278，出现在 $Y'=0.0722$ 。 $B'-Y'$ 分量在黄色($B'-Y'=-0.9278$)时达到其最大负值。类似地， $R'-Y'$ 的极值出现在红和青色时，数值为±0.7874(见图30-1)。这些数值在数字和模拟系统中都不便于使用。将要介绍的 ${}^{709}Y'P_B$ 、 P_R 和 ${}^{709}Y'C_B C_R$ 系统使用的都是经过缩放将分量极值变为更方便数值的 $[Y', B'-Y', R'-Y']$ 版本。

为了从 $R'G'B'$ 获得 $[Y', B'-Y', R'-Y']$ ，对于BT.709亮度系数，使用下面的矩阵公式：

$$\begin{bmatrix} {}^{709}Y' \\ B' - {}^{709}Y' \\ R' - {}^{709}Y' \end{bmatrix} = \begin{bmatrix} 0.2126 & 0.7152 & 0.0722 \\ -0.2126 & -0.7152 & 0.9278 \\ 0.7874 & -0.7152 & -0.0722 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (30-1)$$

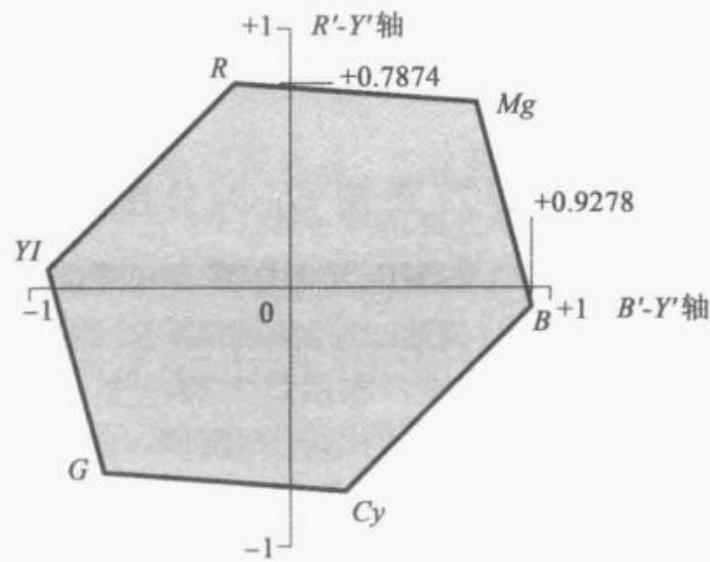


图30-1 BT.709高清 $B'-Y'$ 和 $R'-Y'$ 分量

30.2 BT. 709 高清 $P_B P_R$ 分量

如果要形成与亮度具有相同幅度的两个色差分量，那么要使用 P_B 与 P_R 色差分量。对于 BT. 709 亮度系数，公式为：

$$\begin{aligned} {}^{709}P_B &= \frac{0.5}{1 - 0.0722}(B' - {}^{709}Y') = \frac{1}{1.8556}(B' - {}^{709}Y') \approx 0.5389(B' - {}^{709}Y') \\ {}^{709}P_R &= \frac{0.5}{1 - 0.2126}(R' - {}^{709}Y') = \frac{1}{1.5748}(R' - {}^{709}Y') \approx 0.6350(R' - {}^{709}Y') \end{aligned} \quad (30-2)$$

这些缩放因子将每个色差分量的幅度限制为相对于亮度幅度归一值的±0.5：在上面第一个表达式中 0.0722 是蓝色的亮度系数，在第二个表达式中 0.2126 是红色的亮度系数。在高清模拟接口上，亮度范围从 0mV（黑色）到 700mV（白色）， P_B 和 P_R 模拟分量范围是±350mV。图 30-2 所示的为 $[P_B, P_R]$ 平面图。

用矩阵形式表示，式 (30-1) 的 $B'-Y'$ 和 $R'-Y'$ 行分别用 0.5/0.9278 和 0.5/0.7874 进行缩放。从参考黑电平为 0，参考白电平为归一值的 $R'G'B'$ 进行编码矩阵为：

$$\begin{bmatrix} {}^{709}Y' \\ P_B \\ P_R \end{bmatrix} = \begin{bmatrix} 0.2126 & 0.7152 & 0.0722 \\ -0.114572 & -0.385428 & 0.5 \\ 0.5 & -0.454153 & -0.045847 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (30-3)$$

逆解码矩阵为：

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1.5748 \\ 1 & -0.187324 & -0.468124 \\ 1 & 1.8556 & 0 \end{bmatrix} \cdot \begin{bmatrix} {}^{709}Y' \\ P_B \\ P_R \end{bmatrix} \quad (30-4)$$

30.3 BT. 709 高清 $C_B C_R$ 分量

${}^{709}Y' C_B C_R$ 编码用在分量数字高清设备中。在 8 位系统中，亮度的幅度为 219。色差 C_B 和 C_R 用 8 位偏置二进制形式编码，幅度为±112。高清的 $[C_B, C_R]$ 平面如图 30-3 所示。

在 8 位系统中，在接口上加一个+16 的亮度偏移量，将黑色用代码 16 表示，白色用代码 235 表示；给 C_B 和 C_R 增加+128 的偏移量，所产生的范围为 16~240（公式中的偏置项仍用黑体表示）。高清标准支持 10 位分量，而且 10 位的演播室视频设备很常见。在 10 位接口上，8 位接口的电平和禁用代码保持不变，多出的两位作为最低有效位提供更高的精确度。

从范围是 $[0, \dots, +1]$ 的 $[Y', B'-Y', R'-Y']$ 形成 ${}^{709}Y' C_B C_R$ ，使用以下公式：

$$\begin{array}{l} {}^{709}Y' = 16 + (219 \times {}^{709}Y') \\ 219 \end{array}$$

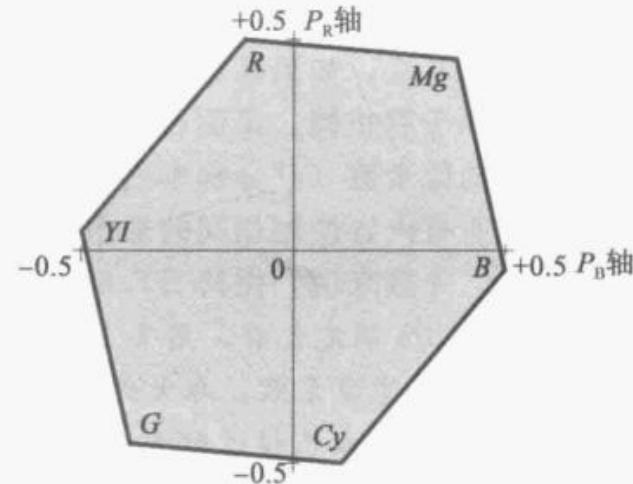


图 30-2 BT. 709 高清 $P_B P_R$ 分量

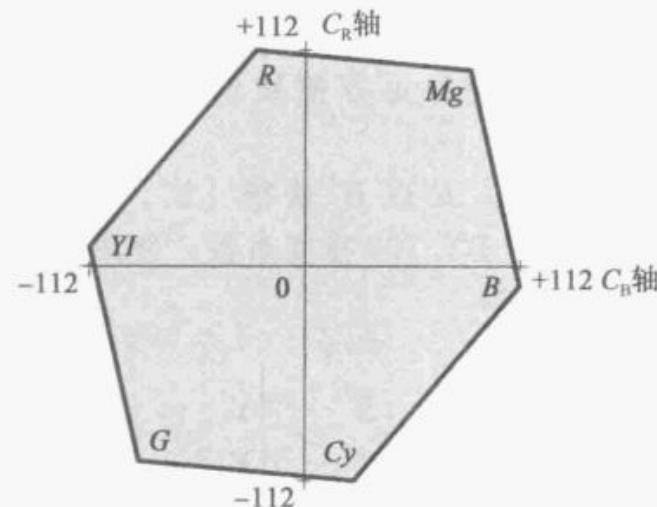


图 30-3 BT. 709 高清 $C_B C_R$ 分量

$$\begin{aligned} C_B &= 128 + \frac{112}{0.9278}(B' - {}^{709}Y') \\ C_R &= 128 + \frac{112}{0.7874}(R' - {}^{709}Y') \end{aligned} \quad (30-5)$$

为了从范围是 0 到 1 的 $R'G'B'$ 得到 ${}^{709}Y'C_B C_R$ ，用因子 [219, 224, 224] 缩放式 (30-3) 中矩阵的各行，缩放因子对应各分量的幅度：

$$\begin{bmatrix} {}^{709}Y' \\ 219 \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 46.559 & 156.629 & 15.812 \\ -25.664 & -86.336 & 112 \\ 112 & -101.730 & -10.270 \end{bmatrix} \cdot \begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} \quad (30-6)$$

矩阵第一行求和得 219，是亮度从黑色到白色的幅度。两个值为 112 的项表示在蓝色和红色时的 $C_B C_R$ 正极值。

从 ${}^{709}Y'C_B C_R$ 恢复范围为 $[0, \dots, +1]$ 的 $R'G'B'$ ，使用式 (30-6) 的逆：

$$\begin{bmatrix} 219R' \\ 219G' \\ 219B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 256 & 0 & 394.150 \\ 256 & -46.885 & -117.165 \\ 256 & 464.430 & 0 \end{bmatrix} \cdot \begin{bmatrix} {}^{709}Y' \\ 219 \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (30-7)$$

${}^{709}Y'C_B C_R$ 分量是 8 位整数；重建的 $R'G'B'$ 缩放到 $[0, \dots, +1]$ 的范围。

图 28-2 显示出当 $R'G'B'$ 分量变换到亮度信号和色差时，单位 $R'G'B'$ 立方体只占包围的立方体体积的一小部分。在数字视频中，只有约 1/4 的 $Y'C_B C_R$ 码字对应 0 到归一值的 $R'G'B'$ 值。某些信号处理操作（如滤波）可能会产生落在合法 RGB 立方体外面的 $Y'C_B C_R$ 码字。这些码字在 $Y'C_B C_R$ 域不会引起困难，但当解码为 $R'G'B'$ 时可能会带来问题。一般来说， $R'G'B'$ 值被钳位到 0 和 1 之间。

30.4 xvYCC 的 $C_B C_R$ 分量

扩展 $R'G'B'$ 系统色域的一种方法是允许分量幅度小于 0 和大于归一值。26.18 节介绍了一种方法。xvYCC 标准是基于 BT. 709 原色的，但允许 RGB 三刺激色值分量的幅度从 $-1/4$ 到 $3/4$ 。

当变换到 BT. 709 $Y'C_B C_R$ 时，所有真实的表面颜色由指针记录，也就是说，所有在指针范围内的颜色产生的值为有效的 $Y'C_B C_R$ 。虽然需要由 BT. 1361 规定宽色域颜色的 $R'G'B'$ 表示，但是通过 ${}^{709}Y'C_B C_R$ 接口传输这些颜色不需要特殊的规定。标记“xvYCC $Y'C_B C_R$ ”或“xvYCC $Y'C_B C_R$ ”明确表示在单位 $R'G'B'$ 立方体外的码字是宽色域颜色，而不应当作非法的 RGB 码字对待。

符合 BT. 1361 的演播室设备尚未部署，预计在几年内也不会有。宽色域采样和制作设备在未来十年左右的时间将开始取代胶片；不过，到那时宽色域消费类显示器还不一定有。当开始部署这些设备时，不太可能它们都有相同的色域；连接到每台显示器的电子产品要根据显示器的属性来处理彩色信号。从长远来看，将不得不采用类似在桌面色彩管理社区使用的色域映射策略。

30.5 从演播室 RGB 到 $Y'C_B C_R$

在演播室设备中，8 位 $R'G'B'$ 分量通常使用和 $Y'C_B C_R$ 亮度分量相同的 219 的幅度。用 8 位二进制算术运算，从范围是 $[0, \dots, 219]$ 的 $R'G'B'$ 编码 $Y'C_B C_R$ ，用因子 256/219 对式 (30-6) 的编码矩阵进行缩放：

$$\begin{bmatrix} {}^{709}_{219} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 54.426 & 183.091 & 18.483 \\ -30.000 & -100.922 & 130.922 \\ 130.922 & -118.918 & -12.005 \end{bmatrix} \cdot \begin{bmatrix} {}^{219} R' \\ {}^{219} G' \\ {}^{219} B' \end{bmatrix} \quad (30-8)$$

用 8 位二进制算术运算, 从 BT. 709 $Y'C_B C_R$ 解码到范围是 $[0, \dots, 219]$ 的 $R'G'B'$:

$$\begin{bmatrix} {}^{219} R' \\ {}^{219} G' \\ {}^{219} B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 256 & 0 & 394.150 \\ 256 & -46.885 & -117.165 \\ 256 & 464.430 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} {}^{709}_{219} Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right) \quad (30-9)$$

30.6 从计算机 RGB 到 $Y'C_B C_R$

在计算机领域, 习惯使用没有上余量和下余量的 8 位 $R'G'B'$ 分量: 黑色是代码 0, 白色是代码 255。用 8 位二进制算术运算, 从范围是 $[0, \dots, 255]$ 的 $R'G'B'$ 编码 $Y'C_B C_R$, 用 256/255 缩放式 (30-6) 的矩阵:

$$\begin{bmatrix} {}^{709}_{219} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256} \begin{bmatrix} 46.742 & 157.243 & 15.874 \\ -25.765 & -86.674 & 112.439 \\ 112.439 & -102.129 & -10.310 \end{bmatrix} \cdot \begin{bmatrix} {}^{255} R' \\ {}^{255} G' \\ {}^{255} B' \end{bmatrix} \quad (30-10)$$

用 8 位二进制算术运算, 从 BT. 709 $Y'C_B C_R$ 解码到范围是 $[0, \dots, 219]$ 的 $R'G'B'$:

$$\begin{bmatrix} {}^{255} R' \\ {}^{255} G' \\ {}^{255} B' \end{bmatrix} = \frac{1}{256} \begin{bmatrix} 298.082 & 0 & 458.942 \\ 298.082 & -54.592 & -136.425 \\ 298.082 & 540.775 & 0 \end{bmatrix} \cdot \left(\begin{bmatrix} {}^{709}_{219} Y' \\ C_B \\ C_R \end{bmatrix} - \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \right) \quad (30-11)$$

30.7 高清和标清之间的转换

在实践中, EBU、SMPTE 和 BT. 709 标准的原色之间的差异可忽略不计。新设备应设计为 BT. 709 标准。另外, 标清和高清实际上已共同采用 BT. 709 规定的传递函数。因此, 用于标清和高清的 $R'G'B'$ 编码在全世界使用本质上相同的参数 (用于桌面计算机的 sRGB 标准使用 BT. 709 原色, 但使用不同的传递函数)。

但是, 正如之前所说, 在标清和高清之间的亮度系数差异显著。假如 HD 系统是孤立的, 这就没什么关系, 但在实践中, 在演播室还是在消费者家中都有标清上转换和高清下转换。如果在转换时没考虑亮度系数的差异, 将会出现严重的色彩重建误差。

原则上, 下转换可以通过先用适当的 3×3 矩阵 (见式 (30-7)) 将 ${}^{709}_{219} Y'C_B C_R$ 解码为 $R'G'B'$, 然后用另一个 3×3 的矩阵 (见式 (29-6)) 将 $R'G'B'$ 编码为 ${}^{601}_{219} Y'C_B C_R$ 来实现。

两个 3×3 矩阵可以结合, 所以转换可一步完成:

$$\begin{bmatrix} {}^{601}_{219} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 1 & 0.099\,312 & 0.191\,700 \\ 0 & 0.989\,854 & -0.110\,653 \\ 0 & -0.072\,453 & 0.983\,398 \end{bmatrix} \cdot \begin{bmatrix} {}^{709}_{219} Y' \\ C_B \\ C_R \end{bmatrix} \quad (30-12)$$

在矩阵的第一行中, 系数 0.099 312 增加了约 1/10 的 BT. 709 C_B 到 BT. 601 的亮度上, 这是 BT. 709 的蓝色亮度系数仅为 0.0722, 而 BT. 601 的蓝色亮度系数为 0.114 的结果。系数 0.1917 增加了约 1/5 的 BT. 709 C_R 到 BT. 601 的亮度上, 这是 BT. 709 的红色亮度系数为 0.2126, 而 BT. 601 的红色亮度系数为 0.299 的结果。显然, 未能进行这种色彩变换将产生很大的色彩误差。

式 (30-12) 和式 (30-13) 没有给出用于亮度的 +16 和用于 C_B 与 C_R 的 +128 的接口偏移量。如果存在偏移, 将其删除, 进行变换, 然后再重新加上。

从标清转换成高清，反转式（30-12）的矩阵为：

$$\begin{bmatrix} \frac{709}{219} Y' \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 1 & -0.115\ 550 & -0.207\ 938 \\ 0 & 1.018\ 640 & 0.114\ 618 \\ 0 & 0.075\ 049 & 1.025\ 327 \end{bmatrix} \cdot \begin{bmatrix} \frac{601}{219} Y' \\ C_B \\ C_R \end{bmatrix} \quad (30-13)$$

不幸的是，对经过如 4：2：2 或 4：2：0 亚采样的图像进行上变换或下变换需要色度插值，颜色变换，然后是色度亚采样，是计算密集型操作。

30.8 彩色编码标准

ITU-R Rec. BT. 709 规定了用于分量模拟高清的 $Y'P_BP_R$ 和用于分量数字高清的 $Y'C_BC_R$ 。1280×720 像素和 1920×1080 像素系统的 $Y'P_BP_R$ 与 $Y'C_BC_R$ 的参数由下面引用的 SMPTE 标准定义。

ITU-R Rec. BT. 709，“用于演播室和国际节目交换的 HDTV 标准的基本参数值”。

SMPTE ST274，“多帧频的 1920×1080 像素扫描和模拟与并行数字接口”。

SMPTE ST296，“1280×720 像素逐行图像采样点结构——模拟和数字表示及模拟接口”。

第 31 章

视频信号处理

本章介绍几个关于视频信号表示和处理的不同主题。

在现代计算机中，用二进制数的 2 的补码形式实现整数算术是很普遍的。当加法或减法这类算术运算结果超出可用的固定位深度时，2 的补码算术一般会出现绕回问题。例如，在 16 位的 2 的补码中，取最大的正数 32 767（或十六进制的 $(7fff)_h$ ），加 1 后得到最小的负数—32 768（或十六进制的 $(8000)_h$ ）。在视频算法的计算机软件实现中，在整数算术运算中允许绕回是一个潜在的问题。在使用整数值的视频信号处理中，必须使用饱和算术。

31.1 边沿处理

如果用一个 25 抽头的 FIR 滤波器（见图 20-26）处理有 720 个样值的一个图像行以产生 720 个输出样值，任何在图像左边沿或右边沿 12 个样值内的输出（结果）将有非零滤波系数与超出图像边沿的输入样值关联。

一种解决这个问题的方法是只产生能从可用的输入样值计算得到的输出样值，在本例中为 696 个。然而，滤波操作经常要级联进行，特别是在演播室，而且在使用一串 FIR 滤波器时反复缩小图像宽度是不能接受的。需要有处理在图像边沿处滤波的策略。

许多数字图像处理（DIP）教科书建议用边沿样值的拷贝来填充像素阵列之外的区域，需要多少次就复制多少次。这种假设对实际上所有成像应用都不现实，因为如果一个小特征碰巧出现在图像的左边沿，复制后，它将变成一个大特征，从而对滤波结果产生过度影响，也就是说，对像素阵列内部产生过度影响。

边沿复制适用于视频压缩中的运动补偿内插：复制的样值用于预测，而不被显示。

有些教科书主张通过取足够多的左边沿样值的镜像来填充图像。在上面的例子中，填充将对最左边的 12 个图像列取镜像。这种方法也是不现实的：在通用成像中，用镜像估计缺少的内容并没有合理的可能性。

许多教科书认为图像包裹在圆筒中：缺少的图像左侧边沿外的样值从图像的右侧复制！这个概念是从傅里叶变换理论得出的，傅里叶变换把有限的数据集视为循环的（周期性的）。这种假设使数学计算简单，但在实践中未经证明，而且这种包裹策略甚至比复制边沿像素更糟。

在视频中，把图像当作是铺在一块黑色区域上来处理：不可用的样值被认为是黑色。用这种策略，反复的低通滤波使隐含的黑色背景在一定程度上侵入图像。在实践中，这种侵入不会引起多少问题。视频图像数据几乎总是包含一些黑色（或消隐）的样值，正如在概述每图像宽度的样值和每有效行的样值时所讲（见 8.3 节）。在演播室标准中，指定位于像素阵列内的一片区域为净窗口，如图 8-4 所示。该区域应当保持在主观感觉上不受图像边沿滤波产生的伪影的影响。

31.2 过渡样值

在 8.3 节中提到，需要避免在一行开头出现从消隐到图像的瞬间过渡的现象。也需要

避免在一行末尾出现从图像到消隐的瞬间过渡的现象。在演播室视频中，一行的开始和最后几个有效的视频样值为消隐过渡样值。建议每行的第一个亮度 (Y') 样值为黑色，而之后的三个过渡样值钳位到满信号幅度的 10%、50% 和 90%。在 4:2:2 时，建议一行的前三个色差 (C) 样值为过渡样值，钳位到 10%、50% 和 90%。图 31-1 显示了过渡样值。过渡值应该用钳位而不是用乘法实现，以避免干扰已经有适当消隐过渡的信号过渡样值。

31.3 图像行

历史上，480i 系统的图像行数没很好地标准化。各种标准指定的“图像行”介于 480 和 487 之间。在 21/284 或更早的行上传输图像是没有意义的，因为在 NTSC 传输中这一行为隐匿字幕数据保留：482 整行再加上底部半行，现在足够了。采用 JPEG、MPEG-1 和 MPEG-2 中的 4:2:0 色度亚采样，要求有 16 的倍数的图像行。现在到处都是基于离散余弦变换 (DCT) 的压缩，以至于在 480i MPEG 视频中 480 行是必需的。在 576i 扫描中，576 行的硬性标准一直是强制的；很偶然地，在 576i 的 MPEG 中，数字 576 恰好是 16 的倍数。

480i 演播室标准在历史上可容纳多达 487 图像行，解释见表 39-1。576i 演播室标准提供 574 全行和两个半行，解释见表 40-1。

MPEG-2 支持 1920×1080 像素的图像格式，但 1080 不是 16 的倍数。在 MPEG-2 编码中，每幅 1920×1080 像素图像的底部填充了 8 个含有黑色的图像行，形成编码的 1920×1088 阵列。多出来的 8 行在解码时被丢弃。

传统上，480i 和 576i 系统的图像阵列包含半行，如图 13-3 和图 13-4 所示：在每帧顶部和底部的行图像信息上加半行消隐。JPEG 和 MPEG 都没有半行消隐：当带半行消隐的图像数据进入 JPEG 或 MPEG 压缩器时，消隐图像数据将被压缩。所幸半行已经从高清中取消了。

演播室视频标准在纵轴上没有过渡样值：意味着从垂直消隐到全画面的瞬间过渡。不过，非图像的垂直间隔信息像视频那样编码，如 VITS 或 VITC，可能会在一帧或一帧中早于图像行出现。有效行仅包含图像行（例外的是在 480i 系统中还包含隐匿字幕数据）。 L_A 不包括垂直间隔行。

有效行（垂直方向）包含图像的高度。有效样点（水平方向）不仅包括图像的宽度，还包括最多达十几个消隐过渡样点。

计算机显示接口标准，如 VESA 的标准，没有对除了消隐之外的非图像（垂直间隔）行做出规定。

31.4 SAL 和 SPW 参数选择

“扫描参数”介绍了两个视频信号参数，每有效行的样点 (SAL) 和每图像宽度的样点 (SPW)。演播室标准中的有效样点数是为了系统设计的方便而选定的；在一个给定的扫描标准中，对应不同采样频率的标准化有效采样点数与采样频率不是精确成正比的。

历史上，对“消隐宽度”而不是图像的宽度进行测量。几十年来，演播室标准和广播标准的消隐宽度有相当大的变化。此外，在不是 50% 的电平上测量消隐宽度导致了对频率响应的依赖性。

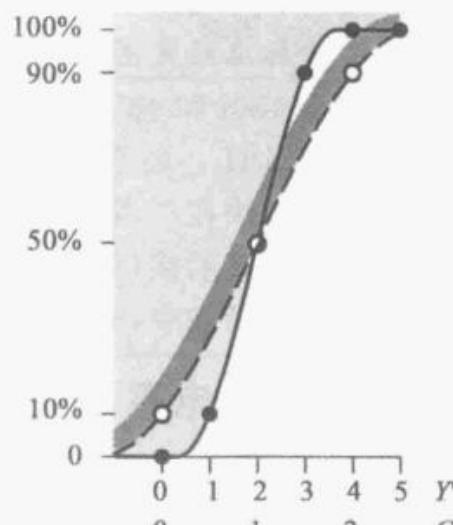


图 31-1 过渡样值

大多数现代视频标准不指定图像宽度：其隐含的意思是，图像在制作窗口内应尽可能宽，还要有合理的消隐过渡。前文的图 13-1 给出了演播室实践的典型 SAL 值。

高清标准规定图像宽度的 50% 点必须位于制作窗口内部不超过 6 个采样点。

对于 480i 和 480p 的数字地面广播，ATSC 认为对过渡样点编码是浪费。ATSC 建立了 704 的 SAL，而不是规定 720 的 SAL。这在制作标准和广播标准之间产生了不一致：MPEG-2 宏块在二者之间是未对齐的。

计算机显示接口标准，如 VESA 的标准，不支持消隐过渡样点，并且没有净窗口的概念。在这些标准中，SPW 和 SAL 是相等的。

31.5 视频电平

前面介绍了 8 位演播室视频电平。演播室视频编码提供了上余量和下余量。在一个 8 位接口上，亮度有代码为 16 的参考黑色和代码为 235 的参考白色；色差用偏移二进制编码，零的代码为 128，负参考电平的代码为 16，正参考电平的代码为 240（讨厌的是亮度和色度的正参考电平不同）。这里用术语参考代替峰值；短暂的幅度峰值有可能超出参考电平。这个范围称为“演播室摆幅”。

所有现代演播室接口可容纳 10 位信号，并且现在大多数设备实现了 10 位。在 10 位系统中，上述的参考电平乘以 4；两个最低有效位提供额外的精度。参考黑色的代码为 64，参考白色的代码为 940。

历史上 480i 系统的视频电平用 IRE 单位表示，有时也简称为单位。IRE 是指美国的无线电工程师协会，IEEE 的前身。参考消隐电平定义为 0 IRE；参考白色电平为 100 IRE。这些值之间的范围是图像幅度。

在模拟接口中，同步编码位比黑色更低的电平；同步是“比黑更黑”。图像幅度与同步幅度之比为图像-同步率。标准化了两种不同的比率：10：4 主要用在 480i 和计算机领域；7：3 在 576i 和高清中通用，偶尔用于 480i，但很少用于计算机领域。

31.6 建立电平（基座）

在北美 480i 复合 NTSC 视频中，参考黑色在消隐之上偏移图像幅度的 7.5% (3/40)。建立电平指的是这个偏移量，表示为图像幅度的分数或百分比。在有建立电平的 480i 系统中，从黑到白有标称 92.5 个 IRE 单位。

在有建立电平的 480i 系统中，术语图像幅度指的是从消隐到白色的范围，尽管严格来说，图像信号的最低电平是 7.5 个单位，而不是 0 个单位。

模拟接口的消隐电平是通过后肩钳位建立的。然而，在带有建立电平的系统中，不存在使接收机能够准确地恢复黑电平的信号元素。如果接口有宽容度差、校准误差、漂移等情况，建立电平会给恢复准确的黑色电平带来问题。因此，建立电平在现代视频系统中已被取消：零建立电平是 EBU N10 分量视频、576i 视频所有变体，以及高清的特性。在所有这些系统中，消隐电平也作为黑色参考电平使用。

在日本 480i 视频起初使用建立电平。不过，在 1987 年左右，零建立电平被采纳；10：4 的图像-同步率被保留。因此，现在标清模拟视频接口有三种电平标准。图 31-2 显示了这些变体。左侧三分之一显示采用 7.5% 建立电平、10：4 图像-同步率的复合 480i 视频的视频电平。这种编码用于一些演播室设备，以及大多数计算机显示接口。中间三分之一显示零建立电平和 10：4 图像-同步率的情况，如在日本使用的 480i 视频。EBU N10 分量视频、576i 系统和高清使用零建立电平、700mV 图像幅度、300mV 同步，如图 31-2

右侧所示。

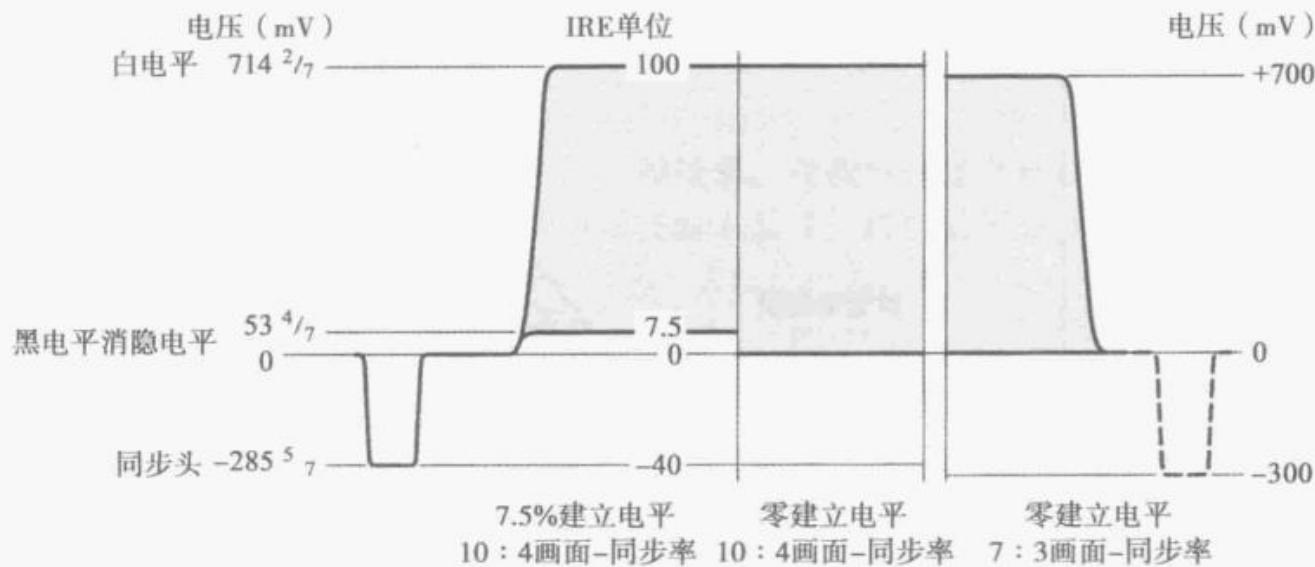


图 31-2 7.5% 建立电平和 0 建立电平的比较

在日本购买的消费类视频设备具有零建立电平。一些日本的消费电子设备的建立电平是可配置为 0/7.5。有时零建立电平有市场术语，如“增强的黑色。”

古老的 EIA RS-343-A 标准规定了单色操作、60.00Hz 场频、2:1 隔行扫描、 $7\mu s$ 的水平消隐和其他在现代视频系统中不存在的参数。不幸的是，大多数 PC 的图形显示标准继承了 RS-343-A 的 10:4 图像-同步率和 7.5% 的建立电平（有些高端工作站具有零建立电平）。

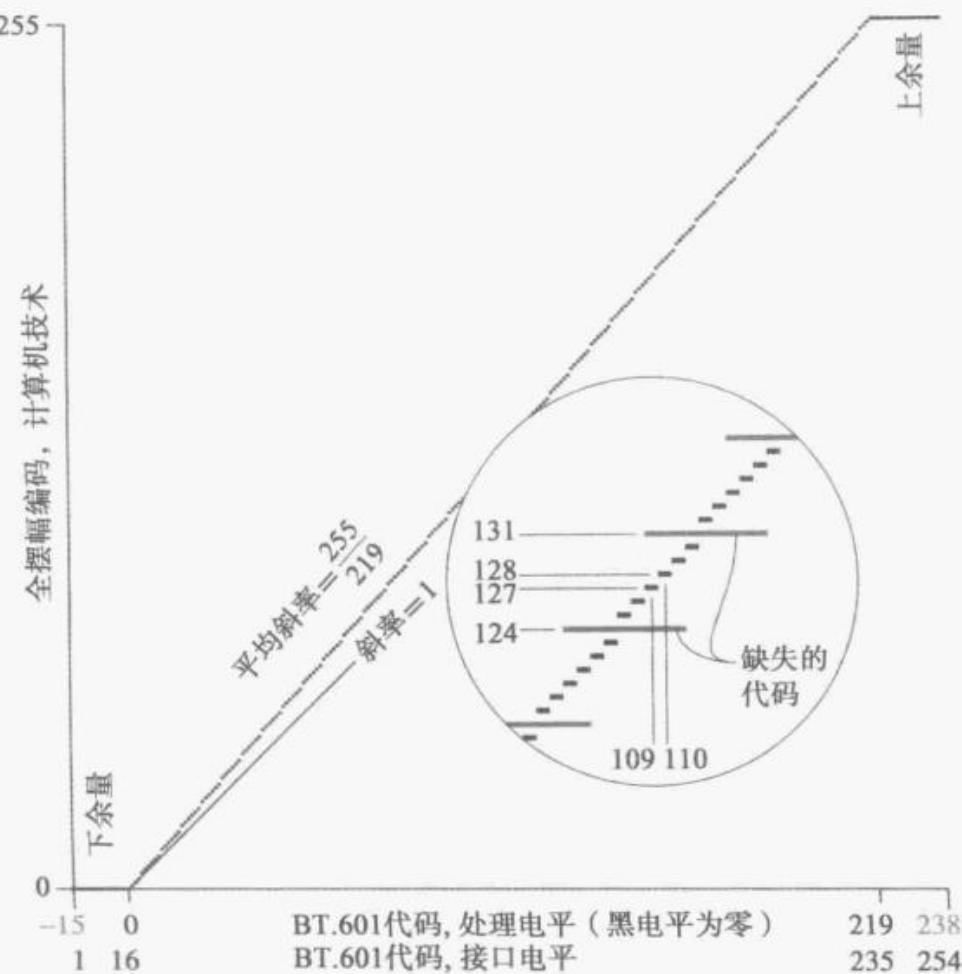
术语基座指的是从消隐电平至黑色电平偏移量的绝对值，用 IRE 单位或毫伏表示：复合 480i NTSC 采用 7.5 IRE 的基座。基座包括任何有意叠加到 R' 、 G' 或 B' 分量，以及亮度和复合视频信号上的偏移量，以达到期望的技术或审美目的。在欧洲，这称为提升（作者认为术语黑色电平比基座或提升更好）。

31.7 从 BT. 601 到计算机领域

计算机图形和演播室视频之间的编码差异需要在接口上进行数据转换。图 31-3 显示了将 8 位 BT. 601 演播室 $R'G'B'$ 转换成计算机 $R'G'B'$ 的传递函数。BT. 601 的下余量和上余量区域被消去，并且输出信号删掉了 36 个代码值。计算机图形和演播室视频之间的编码差异是将演播室视频纳入计算机领域的诸多挑战之一。图中包括乘以一个 $255/219$ (约 1.16) 的比例因子，来处理信号范围的差异。这导致视频信号的上下余量区域被钳位，并导致输出信号丢失几个码值。细节部分显示了在中段的情况；传递函数对称地设置在输入对 [109, 110] 和输出对 [127, 128] 之间。此图显示了从黑到白的线性关系。线性关系适用于计算机系统，在计算机系统中一个斜坡被加载到位于帧缓冲和显示器之间的查找表 (LUT) 里；在这种情况下， $R'G'B'$ 数据显示在计算机显示器上的方式与 $R'G'B'$ 数据显示在视频中类似，见第 27 章。

31.8 增强

本节和随后的几节要讨论增强、中值滤波、核心化、色度转换改善 (CTI) 和扫描速度调制 (SVM)。这些操作表面上都像 FIR 滤波：一个包含一小组邻近样值的“窗口”滑过输入数据。对于每个新的输入样值，滤波操作得到一个输出样值，输出相对于输入有一些固定的时间延迟。与 FIR 滤波不同，除了最温和的增强形式，这些操作都是非线性的。

图 31-3 8 位 BT.601 到全范围（计算机）的 $R'G'B'$ 转换

一般情况下，这些操作无法撤销。

术语“增强”广泛用于图像处理和视频。它没有精确的含义。显然，增强的目的是从某种意义上改善图像的质量。原则上，实现增强只能靠对会降低图像质量的过程的认识。在实践中，极少能得到图像数据经过哪些处理的历史数据，所以不可能有系统的方法来进行增强。

在一些应用中，可能已知图像数据进行过一些引入了特定的劣化或伪影的处理。在这些情况下，增强可能是指为了减少这些劣化而设计的技术。一个常见的例子涉及光圈影响引起的频率响应劣化。在这种情况下的增强，也称为光圈校正，是通过在水平方向或垂直方向或两个方向上进行某种程度的高通滤波来实现的。细节损失 (MTF) 的补偿应在线性光域进行，但有时也在伽马校正域完成。历史上，隔行真空管摄像机（光导摄像管和氧化铅摄像管）的垂直光圈校正是在隔行扫描域完成的。

更一般地，增强会涉及基于对图像数据属性的某些假设的非线性过程。除非信号流控制得非常好，使用这类操作有巨大的危险：当接收到的图像数据没有经过预期的处理，“增强”很容易劣化图像，而不是改善。出于这个原因，作者总的来说非常强烈地反对“增强”。

31.9 中值滤波

中值滤波是一种非线性滤波，其每个输出样值按滤波窗下输入样值的中间值来计算，即其结果是输入值排序后的中间值。通常使用奇数个抽头。中值滤波器往往包括一个 3 抽头的水平窗口；偶尔也使用 5 或 7 抽头。有时使用空间中值滤波器（例如 3×3 ）。

任何孤立的极值，例如冲击噪声产生的巨大样值，从来不会在中值滤波器的输出序列中出现：中值滤波器有助于降低噪声。然而，一个真正的极值也不会被包含在输出序列

中！建议使用中值滤波器要非常谨慎：如果滤波器输入的图像数据其统计特性并非预期，中值滤波很可能劣化图像，而非改善。

31.10 核心化

核心化被广泛认为是用来降低噪声的一种技术。会假设（经常是不正确的）具有低幅度的高频信号分量是噪声。使用互补滤波器将输入信号分离成低频分量和高频分量。低频分量通过，到输出端。估计高频分量的幅度，并对幅度进行阈值操作。如果幅度低于阈值，则丢弃高频分量；否则，高频分量通过，与低频分量求和到输出端。可用图 31-4 所示的框图实现，包括分离低频分量和高频分量的互补滤波器。高频分量由图中的非线性传递函数处理。



图 31-4 核心化电路框图

类似于中值滤波，核心化依赖图像数据的统计特性。如果图像是平面着色的动画，有颜色均匀的大块区域而且区域之间有快速过渡，那么核心化将消除低于一定幅度的噪声。然而，如果输入不是动画，核心化的风险就是会导致它看起来像动画！在面部特写中，皮肤纹理产生低幅度、高频率分量，但它并不是噪声。如果核心化消除了这个分量，面部呈现塑胶般的纹理。

核心化易在图像中引入空间伪影。设想图像中包含一条波斯地毯，地毯在远方消退。地毯的图案在前景（在图像底部）会产生一个相当低的空间频率；随着图案消退在背景中，图案的空间频率变高而幅度变小。如果对这幅图像进行核心化，超过一定距离，核心化会造成图案消失。观看者会感觉到从地毯图案到完全没有图案的突然转变。观看者可能认为超过一定距离有一条不同的地毯或就没有地毯了。

31.11 色度转换改善

彩色下置的录像机表现出很差的色差带宽（证据就是水平方向上差劲的色度分辨率）。亮度的局部性变化可以忠实地再现，但伴随的色差分量的变化将会水平传播。假设彩色区域趋向于均匀地上色，那么一种提高图像质量的方法是检测亮度的局部性变化，并利用该信息实现色差信息的重定位。实现这种方法的技术统称为色度转换改善（CTI）。

使用 CTI 的风险是会带来对边缘的过分强调。另外，CTI 只能用于水平维度：由于造成图像的水平和垂直特征之间能感觉到的（甚至是令人不快的）差异，过度的 CTI 倾向于变得可见。CTI 对动画和某些别的类型的图像效果很好，但是应谨慎使用。

31.12 混合与键控

将视频信号混合产生过渡或分层的效果，例如，混合或擦除，称为合成。在美国，实现这种效果的一件设备（带有控制表面）是一个制作切换台。在欧洲，该设备或操作该设备的人，称为视觉混合器。

在硬件上实现混合、擦除或键控效果需要同步的视频信号，即在垂直和水平域的时序完美匹配的信号。

键控（或合成，或混合）指的是在背景（BG）图像上叠加前景（FG，或填充视频）图像。键控一般由一个键（或蒙版）信号控制；键信号的表示类似亮度，指示伴随的前景

图像数据的不透明度，编码为介于黑色（0，完全透明）和白色（1，完全不透明）之间。在计算机图形学中，键信号（数据）称为 α ，该操作称为合成。

键控（或合成）操作如下式所示。预先乘以键值的前景图像数据在视频中称为成形的，在计算机图形学中称为关联的、积分的或预乘的。未预先乘以键值的前景图像数据在视频中称为未成形的，在计算机图形学中称为未关联的或未预乘的。

$$R = \alpha \cdot FG + (1 - \alpha) \cdot BG$$

键信号有时也称为线性键：修饰语线性并不是指线性光，而是指键信号不是只用完全透明和完全不透明两级表示不透明度。在键控或合成中，上式的合成操作是直接使用的，没有作用于键信号的转换函数。

从历史上看，键控在伽马域实现。然而，对混合的物理特性的正确模拟，需要在线性光域进行键控；现在这样的方法在软件系统中广泛实践（有时通过设置一个选项，标记着类似于“伽马域混合=1.0空间”）；现在越来越多的基于硬件的视频切换器能够进行线性光混合。

键控最困难的部分是提取（“拉出”）蒙版。从计算机图形学的角度来回顾，可参见SMITH, ALVY RAY, and JAMES F. BLINN (1996), “Blue screen matting,” in *Computer Graphics* (Proc. SIGGRAPH), 259-268.

在键控中，前景和背景数据的乘法相当于调制：这会产生高于一半采样速率的信号分量，从而产生混叠成分。为了避免混叠，对前景、背景和键信号上采样；以2倍的视频采样率执行键控操作；然后对结果进行适当的滤波和降采样。大多数键控器直接以视频采样率操作而没有上采样或降采样，因而表现出一些混叠。

图31-5所示的是一个蒙版图像，来自Yung-Yu Chuang和他在华盛顿大学的同事的研究。

为了使合成操作模仿实际场景中的光线混合，键控应在线性光域对前景和背景执行。不过，视频的键控历来在伽马校正域中执行。



图31-5 蒙版图像示例

第32章

帧频、场频、行频和采样率

本章概述 480i 视频、576i 视频，以及高清的场频、帧频、行频和采样率。

用于分量标清的标准采样频率是精确的 13.5MHz。这个速率对 480i 和 576i 都产生每行整数个采样点。有整数帧频的高清标准规定了 13.5MHz 的倍数：720p、1080i30 和 1080p30 系统以 74.25MHz 采样。高清标准还允许以 1000/1001 乘该频率进行操作。

32.1 场频

电视系统一开始使用的场频是基于当地的交流电力线频率：北美用 60Hz，欧洲用 50Hz。

在 20 世纪 40 年代和 50 年代，接收机电源的纹波耦合进电路，比如视频放大器和高压供电，会引起显示照度有规律地波动。如果垂直扫描频率与电力线频率不同，干扰造成的伪影称为哼声（干扰）条，其频率为二者间的频率差，称为拍频。通过选择一个与电力线频率相同的场频可使其可见度最低，以使干扰条是静止的。没有要求说要有精确的频率匹配，或锁定相位：只要干扰图案是静止的，或者漂移很缓慢，就不会令人反感。那些曾造成哼声条的电力线相互作用在现代电路中不复存在，但当时标准化的垂直扫描速率仍然伴随着我们。

将电视扫描锁定到电力线行频的第二个原因关系到图像获取。许多光源，例如荧光灯，以 2 倍于电力线的频率闪烁。如果摄像机工作在与闪烁频率无关的图像速率，则拍摄到的图像容易包含由于闪烁的照明出现的伪影。有各种对策可以克服这些伪影；不过，最简单的防止伪影的方法是以电力线频率的倍数或约数拍摄。

32.2 行频

为 525 行电视选择的光栅线总数是几个小整数的乘积：525 是 $7 \times 5^2 \times 3$ 。选择小的整数因子起因于用真空管分频电路从行频得到场频：这样的分频器只有对小除数是稳定的。每帧的扫描线总数是奇数。等效地，场频为平行频的奇数倍。 $2:1$ 的关系产生 $2:1$ 交织。这些因子结合起来，使单色 525/60 电视的行频为 $30 \times (7 \times 5^2 \times 3)$ ，或精确为 15.750kHz。

对于 525 行接收机，发明了一个方案，用工作在水平扫描频率 15.750kHz 而不是交流线路频率的变压器来产生显像管的高电压。该方法允许使用轻量级变压器，这种变压器称为反激式变压器（该方案今天仍在使用；它可看作开关模式电源的先驱）。反激式变压器是一个复杂的部件，被调谐到水平频率。

当欧洲的工程师开始设计接收机时，把场频固定在 50Hz，以匹配当地的电力线频率成为一种实际需要。欧洲工程师没有从头开发反激式变压器，而是从北美引进！水平频率由此限制在 15.750kHz 附近的一个狭窄范围内。总行数选为 625，也就是 5^4 。单色 625 行电视的行频为 5^6 ，或精确为 15.625kHz；在加入彩色时，行频不变。

32.3 声音副载波

在 1941 年左右，第一届 NTSC 认识到，可以通过使图像行频和伴音副载波频率相干

来使图像中与声音有关的图案可见度最低。在单色 525/60 电视中，伴音副载波被置于行频的 $2000/7$ （即 $285\frac{5}{7}$ ）倍处，即 4.5MHz。传统电视中伴音是调频的，使用的模拟声音调制器即使在静音时也不能保证产生一个精确的 4.5MHz 的调频载波。尽管如此，使调频伴音载波平均为 4.5MHz 也具有一定的价值。

32.4 添加复合彩色

NTSC 和 PAL 彩色编码都采用了频率交织技术实现对单色系统的兼容。采用频率交织，选中的彩色副载波频率可以逐行交替相位，以使编码的彩色在单色接收机上的可见度最低。这种逐行的相位关系使集成了梳状滤波器（虽然可以用更便宜的陷波器来代替）的 NTSC 解码器能够准确地将色度从亮度中分离。

32.5 NTSC 彩色副载波

1953 年，第二届 NTSC 决定选择一个约 3.6MHz 的彩色副载波频率。他们认识到，处理带伴音的复合彩色信号时的任何非线性性，如接收机中频（IF）级的限幅，将导致伴音副载波和彩色副载波之间的互调失真。这两个副载波的差值，或拍频，约 920kHz，落在照度带宽内，可能相当显眼。

NTSC 认识到，如果拍频是行交织的，可使这种图案的可见度最低。由于彩色副载波需为半行频的奇数倍，伴音副载波只好是行频的整数倍。

NTSC 决定彩色副载波应正好是行频的 $455/2$ 倍。拍频的行交织可以通过用分数 $1001/1000$ 把伴音-行频比率（之前为 $285\frac{5}{7}$ ）增加到下一个整数（286）来实现。

美国联邦通信委员会（FCC）的责任曾经是（并且依然是）制定美国的广播标准。假如 FCC 允许伴音副载波频率增加 $1001/1000$ 倍，也就是增加 4.5kHz 至大约 4.5045MHz。假如 FCC 做了这个决定，NTSC 的彩色副载波就正好是 3.583 125MHz；原来的 525/60 行频和场频就不会改变；就可以保持每秒恰好 60 帧并且 NTSC 就不会有丢弃帧！因为伴音是频率调制的，伴音载波反正从来不会晶振稳定于副载波频率（即使在绝对静音的情况下），并且静止频率的容差已经够大（ $\pm 1\text{kHz}$ ）。伴音副载波的偏差曾经而且仍然是 25kHz，所以 4.5kHz 的变化可以用当时的载波差拍声音系统轻松实现。

然而，FCC 拒绝改变伴音副载波，而是通过用分数 $1001/1000$ 把行频和场频降低到分别约 15.734kHz 和 59.94Hz 来满足彩色/伴音限制。彩色副载波确定在 $3.579\overline{545}\text{MHz}$ 。标记为 525/60 的扫描变成了 525/59.94，但不幸的是 525/60 仍被粗心地用于指代 525/59.94。

1001 的因子是 7、11 和 13。这个数值关系在古时候就有名：《一千零一夜》这本书就基于它，数字 7、11 和 13 被认为是非常不吉利的。不幸的是场频 $60/1.001$ ，约 59.94Hz，意味着 60 场占用略多于 1s：每秒计数 30 场不符合时钟时间。丢帧时间码被发明出来以缓解这个难题，见第 33 章。

NTSC 同步发生器历史上采用 $14.318\overline{181}\text{MHz}$ 的主振荡器。这个时钟 4 分频得到彩色副载波，同时 7 分频得到行频的先导。

在 20 世纪 80 年代帧存储同步器出现前，美国的每个主要的广播网络有一个提供 5MHz 的原子钟，后面有一个 $63/22$ 的频率乘法器得到 $14.318\overline{181}\text{MHz}$ 的主时钟。

32.6 576i PAL 彩色副载波

在 576i PAL 中，彩色副载波频率是基于四分之一行频的奇数倍，使用因子 $1135/4$ 。四分之一行频的奇数倍，结合 V 色差分量的逐行相位交替，使 U 和 V 颜色分量占据复合信号频谱的不同部分。逐行倒相使 PAL 信号不受当 NTSC 信号有差分相位失真时的造成的色调错误的影响。

在标准 PAL-B、PAL-G、PAL-H 和 PAL-I 中，基本副载波频率加上了一个 $+25\text{Hz}$ 的偏移量，以使 Hanover 条效应的可见度最低。 25Hz 的偏移量意味着副载波对水平的相位关系每行前进 0.576° 。因此，在 PAL 中的副载波锁定采样不是行锁定的：垂直分布的采样点的副载波相位（ 90° 取模）是不相同的！引入 $+25\text{Hz}$ 偏移破坏了副载波和行频之间的简单整数比率，该比率相当复杂。素因数 64 489 对数字技术来说是相当不可理解的。

32.7 $4f_{sc}$ 采样

最早的数字电视设备对复合 NTSC 或 PAL 视频信号采样。对于复合数字 NTSC 设备来说，以彩色副载波 4 倍的采样频率操作是很方便的，采样频率约为 14.318MHz ，记作 $4f_{sc}$ 。

任何重要的图像处理，例如重定位、缩放、旋转等，要求信号用分量表示。出于这个原因，在制作和后期制作中首选分量视频设备。不过 $4f_{sc}$ 设备有 BT.601 设备一半的数据速率； $4f_{sc}$ 设备比分量设备更便宜，并且统治了标清广播业务许多年。

以 $4f_{sc}$ 对 NTSC 采样，得到 910 个每行采样点 (S_{TL})。总共 768 个样点 (3×2^8) 包含了一行的有效样点，包括消隐过渡。512 (2^9) 行略多于 480i 扫描中的非消隐行数量。数字 768 和 512 对早期存储器系统来说很方便：512 是 2 的 9 次幂，768 是 2 的 8 次幂的 3 倍。在数字电视发展早期，768 和 512 这种组合使帧存储的存储器和寻址电路非常简单。768 和 512 这一特殊组合的重要性现在已经无关紧要了：如今帧存储系统拥有远超过一帧的内存；存储设备有高得多的容量；现在总存储容量是比有效采样点和行数更重要的限制条件。在任何情况下，二进制数 768 和 512 对 576i 帧存储的设计都没有任何帮助。

32.8 常见的采样率

NTSC 和 PAL 系统的设计者基于简单的整数比率选择视频参数。当分量数字采样变得可行时，让人惊喜的是 480i 和 576i 系统的行持续时间之比原来是 144 比 143，见表 32-1 的推导。

对应这些因子的最低公共采样频率 2.25MHz 是现在熟悉的 NTSC 伴音副载波频率 4.5MHz 的一半。 2.25MHz 的任何倍数可用作 480i 和 576i 行锁定采样的基础。最实用的采样频率为 6 乘以 2.25MHz ，即 13.5MHz ；这个乘数是对要求速率足够高以简化模拟抗混叠滤波器设计和要求速率足够低以降低数据传输速率与内存需求的折中选择。

在 13.5MHz ，480i 视频有 858 个每行采样点，而 576i 视频有 864 个每行采样点 (S_{TL})。NTSC 和 PAL 之间的消隐容差支持在两个系统中都是 720 个每有效行采样点 (S_{AL}) 的选择。有效采样点数量的标准化使得标清视频处理设备的设计具有高度的通用性，因为只需要适应有效行计数值的差异就能同时服务于 50Hz 和 60Hz 市场。另外，公

表 32-1 13.5MHz 常见采样率的推导

$f_{H,480i} : f_{H,576i}$
$525 \times \frac{60}{2} \times \frac{1000}{1001} : 625 \times \frac{50}{2}$
$7 \times 5^2 \times 3 \times \frac{5 \times 3 \times 2^2}{2} \times \frac{5^3 \times 2^3}{13 \times 11 \times 7} : 5^4 \times \frac{5^2 \times 2}{2}$
$3 \times 3 \times 2^4 : 13 \times 11$
144 : 143

共的采样频率缓解了标准转换的技术难题，因为无需水平插值。然而，需要区别对待两个系统的消隐来满足演播室交换标准。

32.9 高清扫描的数字学

实际指的是系统中的各个系数。

图 32-1 给出了高清魔法数字的发展过程的图形表示。在左上方为北美的交流电力线频率，还有 525 的因子（所有小整数： $7 \times 5^2 \times 3$ ）。下一个为欧洲的交流电源频率和 625 的因子（也是所有小整数： 5^4 ）。图 32-1 表明在 625/50 系统中有 575 个图像行；这构成每场 287 个整行加一个半行。把每个半行算作一个整行，一共是 576 行。

给 NTSC 系统增加彩色，引入比率 1000/1001，得出 525/59.94 系统。

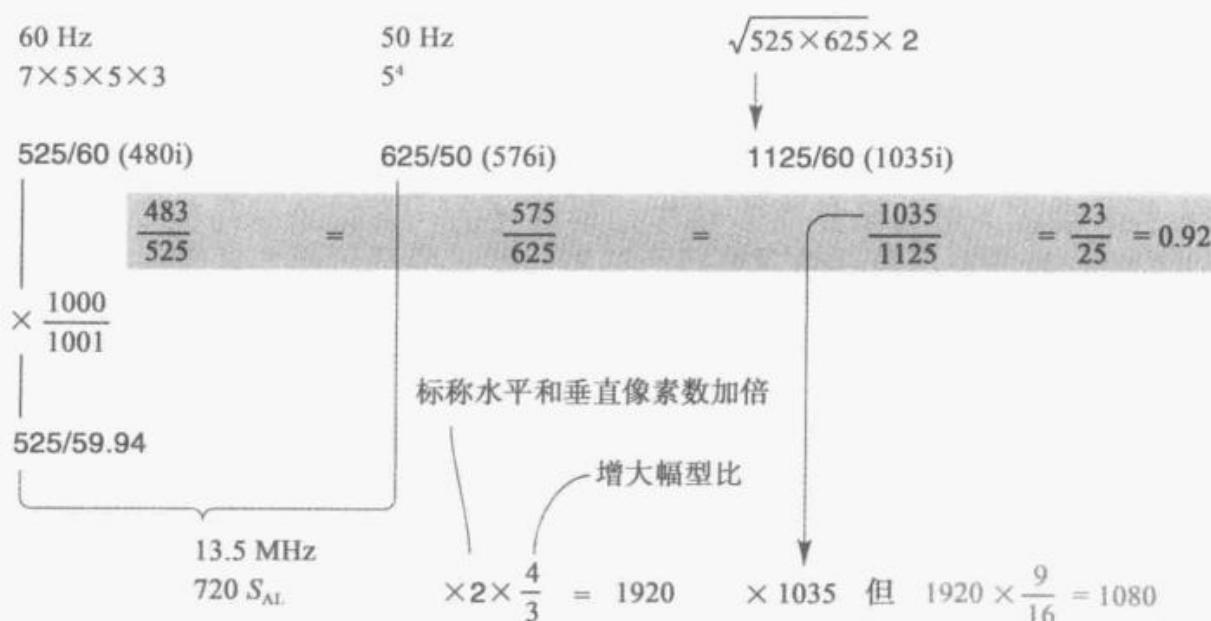


图 32-1 高清扫描的数字学

最初 NHK 设想的高清具有传统电视 2 倍的水平和垂直分辨率：在右上方是高清扫描行总数在概念上的起源，为 525 和 625 几何平均值的 2 倍。北美偏爱 2 倍的 525，而欧洲偏爱 2 倍的 625。设计者选择了总行数为 1125（即 $5^3 \times 3^2$ ），这种折中被认为是大西洋两岸可接受的。

在图 32-1 扫描标记为 525/60、625/50，以及 1125/60 下方，有一个灰色条包含了在每个系统中图像行与总扫描行的比值。这些系统各自的总垂直方向行数 (L_T) 的数量为总行数的 $23/25$ (92%)。这带来了 NHK 在 1125/60 高清中 1035 图像行的原始选择。

对分量数字视频公共采样频率的要求，导致将 480i 和 576i 行频合成为一个公共的采样频率 13.5MHz，以及公共的每有效行采样点 (S_{AL}) 数量 720。对于高清，有效像素数增加 1 倍以提高水平分辨率，然后乘以 $4/3$ 以增加幅型比（从 4 : 3 到 16 : 9），净得 1920。

由这些选择产生了一个维数是 1920×1035 像素的图像阵列，并在 1988 年被 SMPTE 标准化为 240M 标准，却在 1991 年左右，人们已经清楚 1920×1035 像素结构在水平和垂直维度上的采样间距不相等，即不是正方形采样。不相等的程度很小，只有 4%，但对许多应用来说，任何不等距的采样都是负担。在 1995 年左右，标准通过选择图像行数为 1920 的 $9/16$ 倍即 1080 行，来调整为实现正方形采样。SMPTE 和随后的 ATSC 都在 1920×1080 图像阵列中奉行正方形采样。该系统每帧有大约两百万个像素；确切的数目比 2^{21} 略小，正好能存入二进制尺寸的内存器件中。

NHK 计划运行帧频为整 30Hz (“30.00Hz”) 的 1920×1035 像素系统，并且早期的 1035i 设备只工作在这个帧频。然而，1035i30.00 和 480i29.97 之间每 33.367 秒有约一帧时间的差异，这在标准转换中是个很大的麻烦。为了缓解这个问题，并缓解与数字音频采样率相关的工程困难，1080i 高清标准容纳 29.97Hz 和 30Hz 这两种帧频。

当 NHK 和其他机构在开发 1125/60 隔行高清时，其他组织正在开发有着几乎同样像素传输速率的逐行扫描系统，主要是在美国。当时的技术允许大约每秒 6 千万像素的像素传输速率，不论扫描方式是隔行还是逐行。隔行扫描时，在 30Hz 帧频下 60 百万像素/s 允许 200 万像素的图像结构。逐行扫描时，在 60Hz 帧频下 60 百万像素/s 只允许 100 万像素的图像结构。将 100 万像素划分到正方形点阵中就得到了 1280×720 像素的图像结构；这产生了 720p 标准族。

在 2000 年代中期，数字影院社群利用高清设备和基础设施的优势，采纳 1080 图像行用于“2K”标准，却选择了 2048 图像列。这样的选择产生了新的幅型比，约 1.896，之前从未用于影院。本来用 1998 图像列就能实现标准的 1.85 影院幅型比，但显然社群中有些成员担心 1998 不能称为“2K”。选择比 1920 稍大的数字似乎是出于要区分数字影院与高清的（如果不是技术上就是政治上的）短期动机。每边增加的 64 个像素很难说是提高了分辨率。展示的超高清 (Super-HD) 有 2×1920 即 3840 图像列，但“4K”数字影院有 4096 图像列。提出的甚高清 (Ultra-HD) 有 4×1920 即 7680 图像列，但据推测“8K”数字影院将提供 8192 列。在作者看来， $2^k \times 1920$ 的高清相关的图像格式和 2 的幂次的数字影院格式的差异不会持久。高清格式凭借与消费者体量的联系，有可能最终胜出。

32.10 音频速率

数字音频有两个标准采样率：用于专业应用的 48kHz 和用于消费应用的 44.1kHz。在数字音频的标准化中，制造商决定采用两种不同的标准，以区分专业和消费类设备！这个目标悲惨地失败了，而且现在采样率的巨大差别是视频和音频制作中的一个主要的麻烦。

用于消费类数字音频的 44.1kHz 采样率起源于早期的 PCM 录音系统（索尼 PCM-1600），其在 625/50 盒式录像机的 588 个有效行上录制三个（16 位）立体声样值对。在采用 525/59.94 的国家中，原始速率是 $44100/1.001$ ：在每帧的 490 个有效行上录制 3 个 16 位立体声样值对。最终，44.1kHz 的速率在全世界范围内标准化。在 $(1/50)$ s 内，有 882 个样值，为整数。在 $(1/59.94)$ s 内，准确地有 1471.47 个样值：非整数的比率造成大麻烦。

专业的音频采样率选为整 48kHz。每秒 50 场或帧的一幅视频图像的时间间隔对应整 960 个 48kHz 的音频样值。一个 AES/EBU 音频帧包含 192 个左/右样值对，每个样值为 16 位：在 50Hz 视频标准中，一个视频帧正好占用与五个音频帧完全相同的时间间隔。在每秒 59.94 场或帧的视频中，时序关系就没有这么好了。在一幅图像的时间里有 $1601 \frac{3}{5}$ 个音频采样间隔：这个非整数非常不方便。音频样值与视频图像的时序关系每五幅图像才对齐一次。在一幅视频图像的时间里，有 $1001/240$ （即 $4.170\bar{833}$ ）个 AES/EBU 音频帧。

第 33 章

时间码

本章介绍用于视频、电影、录音、编辑和定序设备的时间码的相关技术细节。

33.1 引言

时间码系统为视频的每一帧分配一个编号，允许每个帧被唯一地标识。时间数据采用 HH : MM : SS : FF 形式的 BCD 数字，范围为 00 : 00 : 00 : 00 ~ 23 : 59 : 59 : 29。时间码有用于每秒 24、25、29.97、30 帧的不同版本。时间码数据随着相关联的图像进行数字记录。内嵌时间码 (BITC) 指的是一种将时间码键控加入数字图像内容（即嵌入在图像数据中）的录制方式。

除了 8 位时间数据必需的 32 位外，时间码系统还为每一帧提供了额外的 32 个用户位。用户位可以传送几种类型信息中的一种：第二个时间码数据流（例如原始录制的时间码）；ASCII/ISO 字符流；SMPTE ST 262 规定的运动图像制作数据；辅助 BCD 码数值信息，如磁带卷号；或是非标准的信息。一组 4 个用户位称为一个二进制组。这样，时间码的信息部分一共是每帧 64 位。

时间码的 64 个信息位后附加了一些同步位，以便通过录制通道传输时间码。附加 16 个同步位形成 80 位的线性时间码 (LTC)。附加 18 个同步位和 8 个 CRC 位形成 90 位的垂直间隔时间码 (VITC)，可以被插入视频信号中。

没有 BCD 数字能包含全 1，所以全 1 代码可用于其他目的。有数十个数字的最大可能值小于 8，所以某些时间码数字的高阶位可作为标志使用。这些标志位在表 33.1 中有介绍。

当故意将时间码帧编号的最低有效位锁定到相关联的视频的彩色帧序列时——在 480i 系统中，锁定到 SMPTE 170M 标准的彩色帧 A 和 B，彩色帧标志位被声明。

33.2 丢弃帧时间码

在 25Hz 视频，例如 576i 视频系统，以及 24Hz 电影中，每秒有准确的整数个帧。在这些系统中，时间码与时钟时间有精确的对应关系。

在美国的黑白电视到彩色电视过渡期间，需要满足在水平扫描、伴音和彩色频率之间的某些干扰限制条件。这些限制条件是通过用因子 1000/1001 降低黑白电视的 60.00Hz 场频，产生约 59.94Hz 的彩色 NTSC 场频来解决的。这导致在 29.97Hz 或 59.94Hz 系统中每秒的帧数为非整数。丢弃帧 (DF) 机制可以用来补偿时间码，以获得一个非常接近时钟时间的近似值。当工作在每秒帧数为精确整数时，不需要或不允许使用丢弃帧。丢弃帧时间码在 29.97Hz 或 59.94Hz 系统中是可选的；有直接计数序列的操作称为非丢帧 (NDF)。

29.97Hz NTSC 帧频的计数帧比实际时间慢 1000/1001，非丢帧码将造成每小时约 +3.6s 的明显累积误差。为了使时间码与时钟时间对应，差不多每 1000 帧有一次要丢弃一个帧号，也就是说，从计数序列中删去。当然，丢掉的只是帧号，而不是视频帧！为了维持时间码

(偶数或奇数帧号) 与 NTSC 彩色帧 (A 或 B) 的关系, 帧号是成对丢弃的。

每 $66 \frac{2}{3}$ s——即在 1 分零 6 秒 20 帧的间隔处丢弃一对帧, 将造成丢弃图 33-1 所示的时间码。虽然这个序列不容易看出, 但是它会每整 10min 重复一次! 这是这些数字比率的结果: 2000 个时间码中有两帧, 累积到 18 000 个时间码中有 18 帧, 而 10min 内有 18 000 个 ($1/30$) s 的间隔 (30 帧, 乘以 60 s, 乘以 10 min)。为了产生一个容易计算和记忆的序列, 代替严格的周期性丢弃帧号, 采纳了以下规则: 在每分钟的开始丢弃帧号 00:00 和 00:01, 除了第十分钟。实际上, 该丢弃的一对帧被延迟到下一分钟开始。

图 33-2 给出在时间码显示中形成的表示场识别和丢弃帧代码使用的习惯。

hh:mm:ss:ff
xx:x0:00:00
xx:x1:06:20
xx:x2:13:10
xx:x3:20:00
xx:x4:26:20
xx:x5:33:10
xx:x6:40:00
xx:x7:46:20
xx:x8:53:10

图 33-1 周期性丢弃的时间码

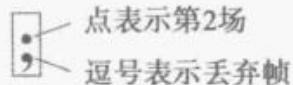


图 33-2 时间码显示

丢弃帧没有实现与时钟时间的完美匹配, 只是达到很好匹配: 以每秒 30/1.001 帧计数丢弃帧代码导致时间码每过 24h 晚 (慢) 约 86.4 ms。如果残差会累积, 11 天或 12 天后时间码约滞后时钟时间 1s。如果时间码序列要维持超过 24h, 时间码应每天在一个适当的时刻停下来被调节到参考时钟时间。没有标准建议这个动作何时发生; 然而, 通常的技术是插入重复的时间码编号 00:00:00; 00 和 00:00:00; 01。编辑设备把重复的代码作为一个时间码中断对待。

33.3 编辑

时间码是视频编辑的基础。一次编辑通过其入点 (待录制的第一场或帧的时间码) 及出点 (录制完成后的第一场或帧的时间码) 来标记。一个编辑过的序列可以用生成这个序列的编辑操作的列表来描述: 编辑决定列表 (EDL) 中的每个条目包含被编辑素材的入点和出点、节目源的入点和出点, 以及磁带盘编号或其他信号源和转换标识。

一个编辑过的节目通常与连续的“非中断”时间码关联。历史上, 编辑设备把 23:59:59:29 和 00:00:00:00 之间的边界作为时间码不连续来处理, 在该点的编辑 (例如开始一个新的节目) 会有问题。因此, 习惯上主节目段以时间码 01:00:00:00 开始。

录像带习惯上在时间码 00:58:30:00 处开始一段 1.5min 的“彩条与单音”先导视频 (在欧洲, 常见在 10:00:00:00 开始一个节目)。

33.4 线性时间码

历史上, 时间码是在演播室录像带和录音带记录器上录制到特性与音轨相似或相同的纵向轨道中, 称为纵向时间码 (LTC)。纵向这个词变得不流行, 而 LTC 更名为线性时间码, 还好它的缩写不变。历史上 LTC 在演播室中作为音频信号对通过三针的 XLR 连接器对接传输。

33.5 垂直间隔时间码

由于固定磁头记录的局限性，VTR 中的纵向时间码不能以非常慢的速度或在磁带停止时读取。垂直间隔时间码（VITC）被数字编码为在垂直间隔中一行或两行上的脉冲流；该方案克服了 LTC 在录像带停止或转动慢时无法读取的缺点。

33.6 时间码结构

表 33-1A 显示了时间码数据的结构。信息位包括一个极性/场标志位（其功能在 LTC 和 VITC 中不同）和三个二进制组标志 BGF_0 、 BGF_1 、 BGF_2 ，在表 33-1B 和表 33-1C 中进行了解释。当 SMPTE 标准调整为适应 25Hz 时间码时，犯了一个拙劣的错误：标志位 BGF_0 、 BGF_2 和极性/场的位置出现了错乱（在该表中用大括号标出）。标志位的解释不幸要取决于时间码是否与 25Hz 有关。时间码信息位没有给出明确的帧频指示；帧频必须在适当的延迟（例如 1s）之后才能确定，或从时间码以外的参数获得。

表 33-1A 时间码位分配表

row	col	7	6	5	4	3	2	1	$\Leftarrow 0$
0	第 1 个二进制组	帧数个位 0~9							
8	第 2 个二进制组（或字符 0）	彩色帧标志		丢弃帧标志		帧数十位 0~2			
16	第 3 个二进制组	秒数个位 0~9							
24	第 4 个二进制组（或字符 1）	极性/场（ BGF_0 ）		秒数十位 0~5					
32	第 5 个二进制组	分钟数个位 0~9							
40	第 6 个二进制组（或字符 2）	$BGF_0 \{BGF_2\}$		分钟数十位 0~5					
48	第 7 个二进制组	小时数个位 0~9							
56	第 8 个二进制组（或字符 3）	$BGF_2 \{\text{极性/场}\}$				小时数十位 0~2			

表 33-1B 时间码标志位

丢弃帧标志	只在 59.94Hz 系统中声明丢弃帧时间码模式
彩色帧标志	当时间码锁定到相关联的视频的彩色帧序列时声明
极性（只针对 LTC）	使全部 80 位 LTC 时间码包含偶数个 0 位（也称为奇偶校验或双相位标记极性校正）
场标记（只针对 VITC）	在第二场时声明

表 33-1C 时间码的二进制组标志

BGF_2	BGF_1	BGF_0	用户位释义
0	0	1	未规定的字符或数据
0	0	1	ISO 646 和 ISO 2202 8 位字符
1	0	1	SMPTE RP 136 数据（“页/行”）
所有其他组合			未分配

在 LTC 中，64 位被串行传送，后面跟着 16 个 LTC 同步位。在 VITC 中，每组 8 位前面有两个 VITC 同步位；这十个字后面跟着最后一对 VITC 同步位和一个 CRC。

LTC 和 VITC 编码的细节见本书的第 1 版。

延伸阅读

用于 480i 电视的时间码在 SMPTE ST 12 中标准化。用于 576i 的时间码在 IEC 60461

中描述，补充信息在欧洲广播联盟文件 EBU Tech. N12 中提供。

SMPTE ST12，“时间和控制码”。

SMPTE RP136 标准化了时间码在每秒 24、25、29.97 或 30 帧的电影胶片上的磁记录。SMPTE ST262 标准化了一种构建用户位数据的方法。

SMPTE RP136，“用于 24、25 或 30 帧每秒运动图像系统的时间和控制码”。

SMPTE 266M 标准化了 VITC 的一个版本，“数字垂直间隔时间码 (DVITC)”，经过 BT.601 4：2：2 接口传输。

SMPTE266M，“4：2：2 数字分量系统——数字垂直间隔时间码”。

SMPTE RP196 标准化了一种用于在 SDI 辅助 (ANC) 数据包内编码时间码数据的机制。

SMPTE RP196，“LTC 和 VITC 数据作为 HANC 数据包在串行数字电视接口的传输”。

第34章

2-3 下拉

电影胶片适用于每秒 24 帧显示。历史上，许多电视节目，包括大量黄金时段的节目，都来自这个帧频的电影（或为 24Hz 的 $1000/1001$ ，即 23.976Hz）。本章讨论将电影转换到高于 24Hz 的帧频。

历史上，电影是用一种称为 2-3 下拉的技术转换为 29.97 帧/s 的视频，其中连续的电影帧先被扫描两次，然后是三次，来产生五个视频场。然后重复进行该过程，把第一和第二场的地位颠倒。该方案如图 34-1 所示。“2-3 下拉”指的是把约 24 帧/s 的电影转换成约 60 场/s 的视频。第一个电影帧转换成两个视频场；第二帧转换成三场。该 2-3 周期重复进行。SMPTE RP 197 将电影帧标记为 A、B、C 和 D。A 帧的特殊性在于正好和同一个视频帧内的两场关联，先是第一场再是第二场（“A 帧”同时表示一个电影帧和一个视频帧）。根据 SMPTE RP201，A 帧时间码的帧计数应该结束于 0 或 5。

除了某些极不寻常的胶片格式，电影胶片垂直地通过摄像机和投影机。小心有人提供类似图 34-1 所示的图，但旋转了 90°。

实时进行电影到视频转换的设备称为电视电影机（术语胶片扫描仪一般意味着非实时操作）。

在 29.97Hz 的系统中，电影播放慢了 0.1%，约为 23.976Hz，因此， $5/2$ 的 2-3 下拉的比例导致场频为 59.94Hz。图 34-1 显示出了 4 个电影帧；该组电影帧旁边的是通过 2-3 下拉生成的视频场序列。右边的标签 1 和 2 表示隔行扫描系统中的第一场和第二场。

2-3 下拉在历史上称为 3-2 下拉。SMPTE 标准将字母 A、B、C、D 分配给一组四个电影帧。A 帧与不包含重复（冗余）场的帧相关联，因此该序列最好被描述为 2-3，而不是 3-2。

当一个 2-3 序列与非丢帧时间码关联时，标准做法是，A 帧使用以 0 和 5 结束的时间码。

在含有 2-3 下拉的序列中，节奏指的是 A 帧的时间规则性。细心的编辑会保留节奏，而粗心的编辑会破坏节奏。

在隔行序列中，场支配指的是在哪一场（第一场或第二场）允许其时间连贯性被编辑操作打断。原则上，在任意场都可进行视频编辑；然而，除了在第一场的开头，在其他任意位置进行编辑都是糟糕的操作。

电影转换到 25Hz 视频是通过把电影加快 4% 播放，把每个电影帧扫描为两个视频场

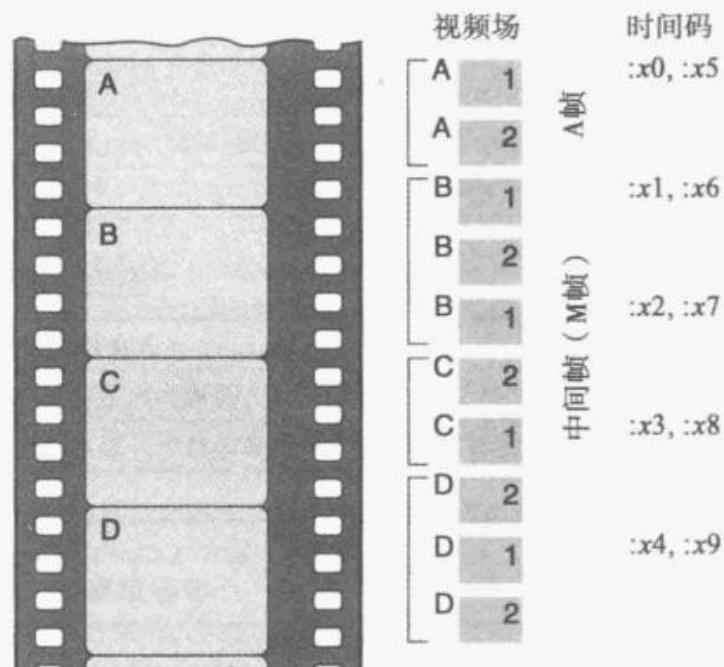


图 34-1 2-3 下拉

的简单方法完成的。可以戏称为 2-2 下拉。2-3 下拉的 0.1% 速度变化对伴音没有显著影响；不过，2-2 下拉的 4% 的速度变化需要进行音频音高校正。

已经介绍了到隔行视频的转换。2-3 下拉技术可用于产生 60Hz 逐行视频。在这种情况下，前面描述的第一场和第二场就是第一帧和第二帧。

工程师容易认为更高的帧频会带来更高质量的运动效果。有提案建议针对电视广播拍摄电影，甚至是以 30 帧/s 而不是 24 帧/s 来拍摄电影。24Hz 的速率显然已是电影的世界标准，也唯一适合转换到 50Hz 系统（通过 2-2 下拉，加快 4%）和 59.94Hz 系统（通过 2-3 下拉，放慢 0.1%）。选择非 24Hz 的电影帧频会妨碍这种已被广泛接受的转换方法，并使电影制片人难以进入国际市场。

34.1 2-3-3-2 下拉

当 23.976 帧/s 的电影进行 2-3 下拉时，视频“中间帧”（时间码 $x2/x7$ ）把电影帧 B 和 C 混合到一起；之后的视频帧把电影帧 C 和 D 混合到一起（包含来自两个电影帧元素的视频帧有时叫作模糊帧）。尽管电影帧 A、B 和 D 可以简单地从单一的视频帧重建，但没有单一的视频帧包含电影帧 C。

这个缺点可用下拉序列的一个小调整来克服，将其表示为 2-3-3-2 下拉，也称为 24pA（24 帧，逐行，高级），如图 34-2 所示。电影帧 C 现在可以从时间码以 3 或 8 结束的视频帧重建。与直接 2-3 下拉相比，采用 2-3-3-2 下拉的视频序列在运动表现上质量略有下降，因为 2-3-3-2 的 6Hz 拍频比 2-3 下拉的 12Hz 拍频看起来更明显一些。然而，该方案对于桌面视频编辑很流行。

2-3-3-2 下拉被一些制造商标记为 24pA（表示高级的）。当与 DV 录制一起使用时，2-3-3-2 的优势是可以不解压视频位流就能重建电影帧。

“2-3-3-2 下拉”与 2-3 下拉是相同的，除了第 8 个视频场表示电影帧 C 而不是 D。这种方法相较于 2-3 下拉的优势在于电影帧可以从完整的视频帧中重建，而不需要“拼接”两场。35mm 电影胶片垂直通过摄像机和投影机。该方案也称为 24pA。它不用于广播。然而，如果用于广播，运动损伤也不比 2-3 下拉差很多。

34.2 将电影转换到不同的帧频

当来自 2-3 下拉的图像序列显示在视频中时，运动表现会受一定程度的损伤。这种损伤很少会令人反感。然而，如果一个 2-3 序列简单地转换到不同的帧频，或者从 2-3 序列中提取一个静止帧，那么所造成的损伤容易令人反感。在对电影原件进行帧频转换之前，需要对原始的电影帧进行重建，用一个称为逆电视电影或逆 2-3 下拉的过程，将合适的场对交织在一起。尽管 2-3 下拉已经使用了半个世纪，但没有辅助这种“交织”的信息与视频信号一起提供。

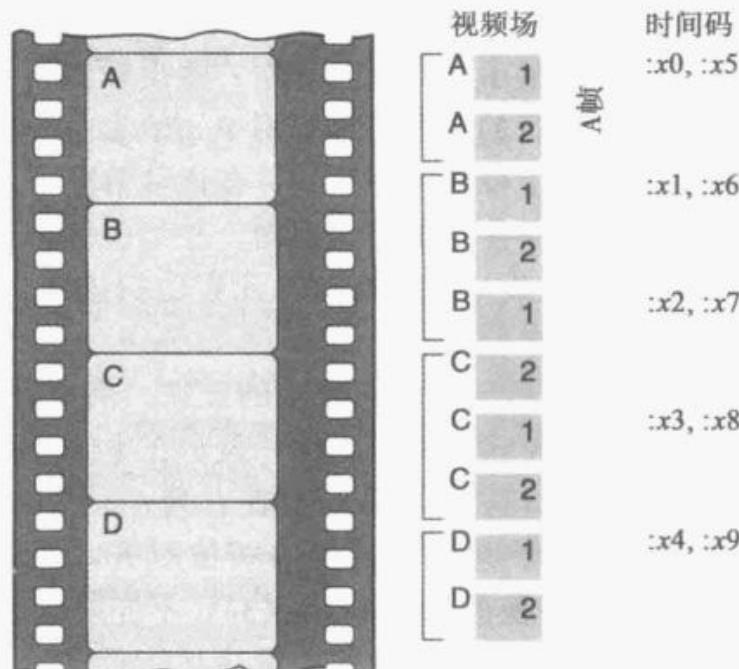


图 34-2 2-3-3-2 下拉

一种可以将 24Hz 电影帧频转换为任意扫描速率的简单方法是把连续的电影行以电影扫描速率写入一个双端口帧缓冲器，然后以视频扫描速率从缓冲区读取出连续行。不过，如果一个场景元素是相对于相机运动的，这种技术将不起作用。图 34-3 右边部分表示从电影扫描的行正被写入帧缓冲区。斜虚线与视频扫描交叉；在线相交的垂直坐标处，产生的画面突然从一场切换到另一场。这使得输出场包含的空间不连续（尽管这个描述是指隔行扫描，但这些效果都与隔行没有直接关系；在逐行系统的 2-3 下拉中也有同样的效果）。

图 34-3 显示了 2-3 下拉的垂直-时间 ($V \cdot T$) 关系，横轴表示时间，纵轴表示扫描的垂直维度。虚线表示电影采样，实线表示视频采样。电影采样在同一时刻对整幅图像进行采样。2-3 下拉引入的交错序列是电影采样线的间距不规则的原因。在视频中，采样延迟到扫描进行到的那一场。这反映为图中线的倾斜。

在专用显示器的去隔行中，输出帧频可以锁定到 59.94Hz 或 50Hz 的输入视频场频。然而在去隔行的桌面计算应用中，输出一般高于 60Hz 且与视频速率不同步：无法强制输出速率与原生视频速率匹配。在这种情况下，从存储器逐行读取的行为快于并且异步于隔行写入。如果使用单个帧存储，在某些时刻“快的”读指针将越过“慢的”写指针。如果“指针交叉”发生在含有运动元素的扫描线上，将在图像中引入空间干扰。这些干扰可通过使用三场存储来避免。

图 34-4 显示了空间域中的效果。两个完整的电影帧显示在左侧。2-3 下拉技术在视频序列中引入时间的不规则性，显示在中间一列的 5 个视频场，但单个图像都是完整的。简单的帧缓冲方法的效果显示在右列：5 场中有两场产生了空间不连续。从 24Hz 转换到 60Hz，依赖于电影和视频的相位对齐，两处或三处不连续对于观看者都是明显的。将 24Hz 电影转换为 59.94Hz 视频，不连续的效果会缓慢向屏幕下方漂移。

使用一对缓冲器，即双缓冲，并用电影和视频帧的起始来同步缓冲器的写入和读出，可以保持每一帧的完整并消除空间上的不连续。不过，这项技术涉及的延时又引入了与 2-3 下拉相同的时间不连续！

虽然 2-3 下拉在视频流中引入了时域伪影，但是当以视频速率显示 2-3 视频时，能得到可以接受的运动表现。然而，如果再一次改变帧频，两个级联转换的时域伪影会变得令人反感，尤其是用宽画面角度显示图像时。

在经过小心完成的电影转换视频中，以时间码数字 0 或 5 结束的视频图像对应电影源的 A 帧。除了与 A 帧锁定的时间码外，没有与视频信号一起传输关于电影来源信息的标准。没有锁定的时间码，检测 2-3 下拉的唯一方法就是在连续场中进行数据比较。如果两个连续的第一场在一定的噪声容限内包含相同的亮度和色度，并且重复模式符合 2-3 序列的特性，则可以假定该素材是来自电影的。一旦识别出原始的电影帧，实现到最终显示速率的转换就可以引入最少的运动伪影。在专用硬件上用这种方法识别电影帧是可行的，但难以在现在的台式计算机上实现。

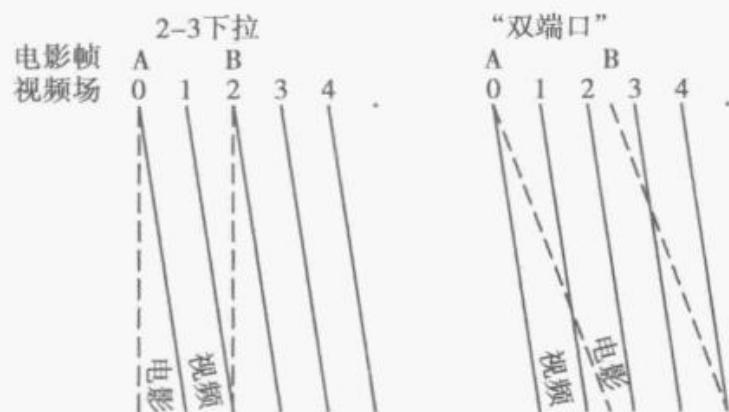


图 34-3 2-3 下拉的垂直/时间关系

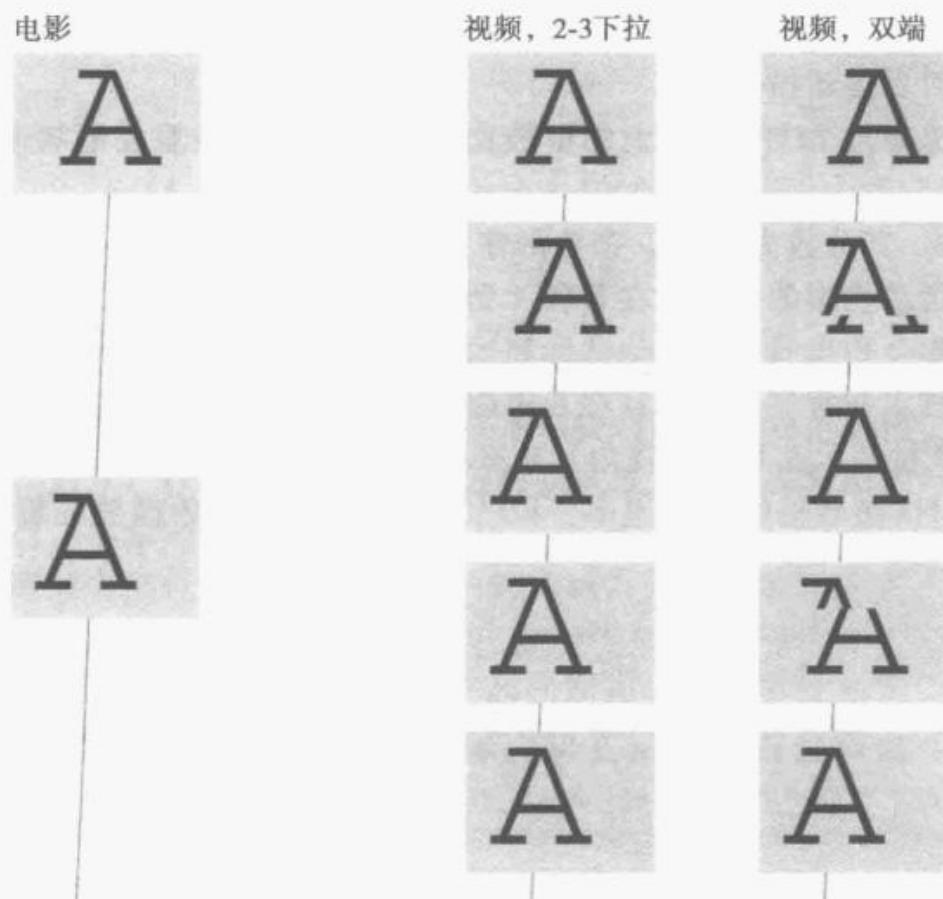


图 34-4 2-3 下拉的空间视图

34.3 原生 24Hz 编码

传统上，2-3 下拉在演播室进行，在将电影转换为视频时使用。重复的场是冗余的，并且除了兼容原生 60 场/s 的设备，消耗介质容量还没有好的效果：当把电影记录在 VHS 磁带等介质上时，整整有 20% 的介质容量被浪费了。在演播室中，难以从 2-3 序列中恢复出原始的电影帧。重复场的信息无法直接获得，解码设备常常通过在连续的场中比较像素值来试图重建序列。

MPEG-2 编码可处理对 24 帧/s 的逐行素材的编码；编码一个 2-3 下拉序列的效率较低。一些 MPEG-2 编码器配备了检测和删除 2-3 下拉的功能，所以能编码每秒 24 个逐行帧。然而，这一过程称为逆电视电影，很复杂并且容易发生问题。

尽管 MPEG-2 本身允许 23.967Hz 或 24Hz 原生帧频，但 DVD 不允许。480i DVD 上的电影素材编码为 59.94Hz 隔行。解码器通过标志位识别电影帧，然后插入 2-3 下拉输出 480i29.97，或按照相应的配置输出 480p23.976，见 47.3 节。

通常，会从一个 MPEG-2 位流中删除 2-3 下拉的重复场。不过，位流可以包括标志，通知解码器在显示时应该重复被删除的场，有了这些标志，解码器可以重建 2-3 序列用于 59.94Hz 显示。或者，当配置为逐行输出时，解码器可以重建 23.976Hz 的帧，并直接在合适的显示设备上显示出来，或 3 倍帧频到 72Hz 输出（实际上是 72/1.001 或约 71.928Hz）。蓝光就避免了所有这些复杂操作，可以直接编码 23.976Hz 逐行帧。

34.4 转换到其他帧频

在从 24Hz 到 60Hz 的 2-3 下拉中，连续电影帧的信息在固定的序列 {2, 3, 2, 3, 2, 3, …} 中重复。重复图案的频率（即拍频）是比较高的：{2, 3} 图案每秒重复 12 次，所以拍频为 12Hz。在这种情况下的帧频比例是 5/2，分数中小的整数值决定了高拍频。

当 PC 从如 DVD 或数字卫星的节目源显示来自电影的 29.97Hz 视频时，情况更为复杂，并且更可能引入运动损伤。

当待转换的速率与 24Hz 有更大的整数比例关系时，图案重复的频率会下降。例如，转换为 75Hz，其分数为 $25/8$ ，产生的序列为 {3, 3, 3, 3, 3, 3, 3, 4}，每秒重复 3 次；转换为 76Hz，其分数为 $17/6$ ，产生的序列为 {2, 3, 3, 3, 3, 3}，每秒重复 4 次。人类视觉系统对运动伪影的敏感性在每秒 4 到 6 拍之间达到峰值；在转换为 75Hz 或 76Hz 时引入的时间伪影会相当明显。运动估计和运动补偿插值有可能用于减轻这些转换伪影的严重性，但这些技术非常复杂，并且在几年内对于桌面计算仍是不可企及的。

当 24Hz 的素材显示在具有宽视角、良好周边条件的计算环境时，减少运动伪影的最好办法是选择 24Hz 整数倍的显示速率。以 60Hz 显示重现视频播放环境，运动表现很易于被人接受。

第 35 章

去隔行

8.4 节解释了当隔行扫描摄像机拍摄下相对于摄像机运动的场景元素时，场景元素会出现在两场的不同位置。逐行帧的重建对某些图像处理操作，如上转换、下转换或标准的转换，是必要的。另外，计算机图像通常用逐行格式表示：将视频图像与计算机显示器相结合也需要去隔行。本章概述去隔行技术。

首先介绍空间域去隔行，然后描述垂直-时间域，并概述实际的去隔行算法。

参考图 35-1 中的测试场景，讨论去隔行的问题和算法。测试场景包括一个黑色的背景，部分被一个相对于摄像机在运动的白色圆盘遮挡。



图 35-1 测试场景

35.1 空间域

视频每秒拍摄 50 或 60 个不同的场。如果场景中包含一个相对于摄像机运动的物体，每场将携带该物体一半的空间信息，但在第二场中的信息将根据物体的运动而发生位移。图 35-2 说明了这种情况，分别显示了第一场和第二场。这个例子是典型的由电荷耦合器件 (CCD) 摄像机用很短的曝光时间拍摄的场景；这个例子忽略了由于相机曝光时间不为零而产生的拍摄模糊（关于图像获取和显示的时域特性的详细信息，见作者的 SMPTE 论文）。

与逐行信号相比，可以把隔行视频信号看作行的重排。一种明显的实现去隔行的方法是，按视频顺序写入两场视频存储器，先是第一场，之后是第二场，然后逐行读出组合在一起的帧（以空间顺序）。这种方法有时有一个复杂的名字，称为场复制或交织。这种方法非常适合静止场景，或只包含缓慢移动元素的场景。然而，第二场的图像数据相对第一场的图像数据延迟了半帧的时间（通常为 $(1/60)$ s 或 $(1/50)$ s）。如果场景中包含了运动较快的元素（如测试场景中的圆盘），并且认为图像数据属于同一时刻，将引入场撕裂（field tearing）：无论是作为静止帧还是作为运动画面观看，重建的场景元素将带有锯齿状边缘，效果如图 35-3 所示。

通过场内处理可避免场撕裂，场内处理只使用来自视频单个场的信息。最简单的场内技术是在逐行读出时复制每一行。一个缺点是，这种方法重现静止元素时的垂直分辨率最多只有其可能的垂直分辨率的一半。另外，行复制在画面中引入了结块（blockiness），并有一个图像行的明显下移，效果如图 35-4 所示。

通过合成明显在空间上位于相对场但在时间上位于相同场的信息，可避免行复制方法产生的结块。可以通过平均一个场内垂直方向上相邻的样值的方法来实现，得到一个合成的中间行，如图 35-5 所示（在计算机行业，这种方法称为“bob”）。求平均可以在写入视频存储器之前，或者在读取时，取决



第一场



第二场

图 35-2 隔行采集



图 35-3 交织技术



图 35-4 行复制

于哪种方法对存储器系统更有效。求平均的方法减轻了结块的缺点，但不能补偿垂直分辨率的损失。然而，该方法对 VHS 级别图像的处理效果很好，这种图像在任何情况下都缺少分辨率。与其简单地对两行求平均，使用更长的具有合适的抽头权重的 FIR 滤波器能提高性能，见第 20 章。

为了适度提高两抽头平均的效果，使用系数为 $[1/16, 7/16, 7/16, 1/16]$ 的四抽头滤波器。



图 35-5 场间平均

可以在垂直-时间 ($V \cdot T$) 域考虑隔行到逐行的转换。图 35-6 显示了图 35-2 所示的隔行拍摄场的三维视图。从“侧面”看，沿着扫描线的轴线投影出垂直-时间域。时域采样点位于对应场时刻的离散时间点上；垂直采样点位于由扫描线间距确定的空间离散间隔上。图 35-6 所示的四个空心圆圈表示在某一场时刻某一行上可用的原始图像信息的采样点。对这些样值的计算可以合成图案中心缺少的采样值。在下一幅图中，重建的样值将绘制成实心的圆圈（在给定的垂直和时间坐标上，沿着扫描线对每个采样点进行类似的计算：对于 BT. 601 数字视频，每个扫描行进行 720 次计算）。

图 35-7 以二维视图显示了垂直-时间域。从隔行到逐行的转换需要计算以空圆圈表示的四个采样值的某种组合，来合成其中间的样值（用实心圆圈表示）。使用多于 4 个样值的技术是可行的，但对桌面视频来说复杂度过高。

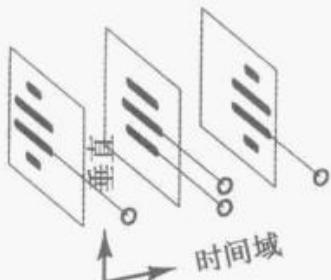
图 35-6 $V \cdot T$ 形成图 35-7 $V \cdot T$ 域

图 35-8 在 $V \cdot T$ 域中显示了场复制（或交织）技术。待计算的样值简单地从前一场复制。结果在空间上是正确的，但如果图像的对应区域包含运动的元素，将会引入图 35-3 所示的撕裂。

代替从前一场复制信息，可以对前一场和后一场求平均。图 35-9 显示了这种方法。这种技术也会有某种形式的场撕裂，但它与一种马上要讨论的自适应方法配合会很有用。

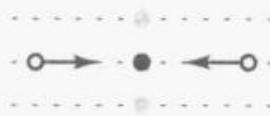
图 35-8 $V \cdot T$ 域内的静态点阵（交织）图 35-9 $V \cdot T$ 域内的帧间平均

图 35-10 在 $V \cdot T$ 域中显示了行复制技术。中间样点简单地从上面一行复制。因为被复制的样点来自同一场，所以没有引入时间伪影。行复制技术会造成向下移动一个图像行。从图 35-4 看，这种移位很明显：测试场景中的圆盘是垂直居中的，但在图 35-4 中显得偏离了中心。

图 35-11 显示了场内平均或称为 bob 技术。中央样点是对期望位置之上一行和之下一行的样点求平均得到的。因为被求平均的信息来自同一时刻，所以没有引入时间伪影。此

外，避免了行复制带来的向下移动一行。然而，静态场景的垂直分辨率降低。

图 35-10 $V \cdot T$ 域内的行复制图 35-11 $V \cdot T$ 域内的场内平均

BBC 的 Martin Weston 发现，使用两场和四行存储，不需要自适应，采用精心选择的系数可能实现很好的去隔行效果。Weston 的滤波器系数如表 35-1 所示，阴影单元格对应滤波结果。

表 35-1 Weston 去隔行器

图像行	场 $t-1$	场 t	场 $t+1$
$i-4$	32		32
$i-3$		-27	
$i-2$	-119		-119
$i-1$		539	
i	174	*	174
$i+1$		539	
$i+1$	-119		-119
$i+1$		-27	
$i+1$	32		32

35.3 运动自适应性

在 $V \cdot T$ 域中分析转换过程表明，可以通过用静态方法转换静止的场景元素，而用行平均转换运动的元素来改善性能。这种改善可以通过检测每个要得到的像素是否可能属于某个运动中的场景元素来实现。如果该元素可能是在运动中，则使用场内平均（避免空间伪影）。如果该元素可能是静止的，则使用场间平均（避免分辨率损失）。

通过将一场与前一场比较可检测出运动。理想情况下，将会用到一个相似场：如果要估计场 1 的运动，那么之前的场 1 应当用作一个参照点。然而，这种方法要求有一个完整的帧存储器用于运动检测。根据具体的应用，使用单场存储器从相对场检测运动可能就足够了。

无论是用一场存储器还是一帧存储器来进行运动检测，重要的是用空间低通滤波器对可用的图像信息进行过滤，以防止小细节或噪声造成运动估计的突然变化。图 35-12 显示了用于计算扫描行之间空间采样值的空间低通滤波器系数。阴影方块表示结果的有效位置。该滤波器需要一个行存储器（或双端口存储器）。加权和可以通过三个级联的 $[1, 1]$ 段来实现，每个段都需要一个加法器。

与一个扫描行同位（空间上重合）的低通滤波器的样值可以用图 35-13 所示的权重计算得到。带阴影的方块表示要进行运动检测的中央样点。只用行存储和级联的 $[1, 1]$ 段也可实现该滤波器。用两个空间滤波结果的差的绝对值估计运动的概率。

图 35-12 和图 35-13 所示的空间滤波器集成了系数为 $[1, 4, 6, 4, 1]$ 的横向滤波器。这些特定的系数能够用级联的 $[1, 1]$ 滤波器实现。图 35-12 所示的 2 行空间滤波器可以用一个行存储器、两个 $[1, 4, 6, 4, 1]$ 横向滤波器和一个加法器来实现。图 35-13 所示的 3 行空间滤波器可以使用两个行存储器、三个 $[1, 4, 6, 4, 1]$ 横向滤波器和两个加法器来实现，其中一个横向滤波器的结果要乘以 2 来实现系数 2、8、12、8、2。

当运动估计超过某个阈值时，一种简单的自适应滤波器会从帧间平均切换到场间平均。然而，突然的切换会造成伪影：两个相邻采样点可能具有非常相近的值，但如果一个被判定是静止的而另一个被判定是运动的，由去隔行滤波器计算出的样值可能有很大差异。这些差异可能在视觉上让人感到反感。可以通过按比例混合（即渐变）帧间和场间平均值而不是突然切换来减少这些伪影。混合可由一个运动差值的窗函数来控制，如图 35-14 所示。



图 35-12 间置性空间滤波器系数



图 35-13 同位空间滤波器系数

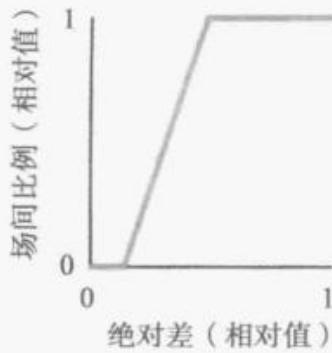


图 35-14 去隔行的窗函数

延伸阅读

Bellers 和 de Haan 写了关于去隔行技术的权威书籍。该书着重于 Philips 的专利技术并见于 NXP 的 VLSI。关于去隔行技术的概要可在 Madisetti 的书中“DEHAAN 和 BRASPENNING”一章找到。

BELLERS, ERWIN B. and DE HAAN, GERARD (2000), *De-inter-lacing: A key technology for scan rate conversion* (Elsevier/North-Holland) .

DE HAAN, GERARD and BRASPENNING, RALPH (2010), “Video Scanning Format Conversion and Motion Estimation,” in MADISSETTI, VIJAY K., *The digital signal processing handbook*, Second edition, Vol. 2 (Boca Raton, Fla. , U. S. A. : CRC Press/Taylor & Francis) .

第 36 章

彩条

36.1 标清彩条

图 36-1 所示的为经典的 SMPTE 彩条测试图案生成的图像。图像的上部 2/3 包含 100% 白条，其后是 75% 饱和度的基色和次级色彩。中部窄带包含“反色彩条”(reverse bars)；这一段可以设置复合 NTSC 或 PAL 解码器的色调和色度。图像的底部 1/4 为相位为 $-I$ 的副载波、白色条和相位为 $+Q$ 的副载波，以及（右侧的）PLUGE 元素，稍后将做介绍。

ANSI/EIA-189-A，编码彩条信号（原标记为 EIA RS-189-A）。

图 36-2 示出了上部 2/3 图像中 SMPTE 彩条的 $R'G'B'$ 分量。每个扫描行是由一个具有零值或归一值的红、绿、蓝二进制序列产生的。各分量按对亮度的贡献大小排序，使得最终的亮度分量由左到右递减（SMPTE 彩条下方 2/3 处的窄带已倒置了计数顺序并且其绿色分量强制为零）。

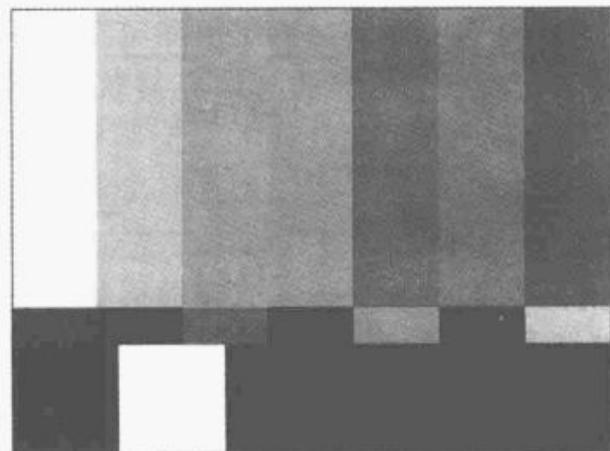


图 36-1 SMPTE EG 1 标清彩条测试信号

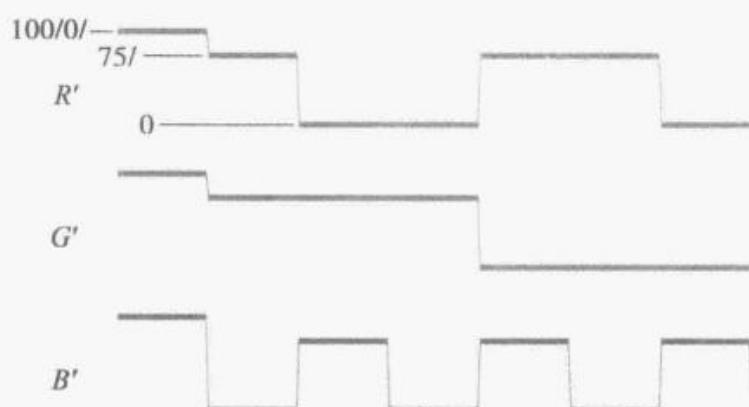


图 36-2 彩条的 $R'G'B'$ 基色分量

SMPTE 彩条 $R'G'B'$ 基色分量幅度为 75 IRE，记作 75/0/75/0。图 36-2 所示为 100/0/75/0 彩条，白色条幅度为 100 IRE。

在演播室设备、分量视频和 PAL 广播中，信号的处理、记录和传输信道可以容纳范围从 0 到 1 的 $R'G'B'$ 分量组合而成的各种编码信号。100% 彩条信号包含了在这些极限值上的 8 个点。

完全饱和的黄色和完全饱和的青色可以使得复合 PAL 信号的峰值达到基准白色的 $4/3$ ($133\frac{1}{3}\%$)。然而，NTSC 发射机的复合信号幅度限制在基准白色的 120%。如果直接向 NTSC 发射机中输入一个 100% 色条，输出信号会被限幅。为避免限幅，通常用 75% 的色条测试 NTSC 传输。白色条由幅度为 100% 的三基色组成，而其他色条的三基色幅度减小至 75%，以便将复合 NTSC 彩条的峰值限制在参考白电平内。虽然“75% 色条”的习惯出于充分的理由被采纳，但由于模拟 NTSC 发射机已经退役，沿用该习惯不再有充分理由。

100%彩条的 $R'G'B'$ 分量的 $R'G'B'$ (或 RGB) 取值为 0 或归一化值，独立于基色的色度：由于基色不同，SMPTE、EBU 和高清标准下的彩条信号并不完全相同。

36.2 标清彩条标记法

前文介绍了 100% 和 75% 彩条，在使用的还有另外几个变体。存在如此多种彩条，以至于要有国际标准来进行标记。一个彩条信号用被斜杠分隔的四个数字来表示，每个数字为单位数（原为 IRE 单位）。第一对数字分别给出无色条，即黑色条或白色条的三基色的最大值和最小值。第二对数字分别给出彩色条的最大值和最小值。

见 ITU-R Rec. BT. 471《彩条信号的命名原则与描述》。

前面描述的 100% 彩条记为 100/0/100/0。该变体应用于演播室、所有形式的分量视频和 PAL 制式的传输。在使用 7.5% 建立电平的老式 480i 复合 NTSC 系统中，100% 彩条指 100/7.5/100/7.5。该变体曾在演播室中使用。然而如前文所述，地面模拟 NTSC 传输无法处理 100% 彩条。NTSC 制发射机使用带有建立电平的 75% 彩条进行测试，记为 100/7.5/75/7.5。日本采用无建立电平的 480i 视频，也使用 75% 彩条，记为 100/0/75/0。

36.3 PLUGE 元素

彩条图案的右下象限是画面行提升产生设备 (PLUGE) 生成的元素，如图 36-3 所示。其缩写起源于“产生设备”，但现在的 PLUGE 符号代表信号元素。叠加在参考黑电平上的是两个元素，一个略低于参考黑电平，另一个略高于参考黑电平。

彩条信号的 PLUGE 元素允许精确地设置黑电平。576i PLUGE 中 14mV 的行程相当于 ± 2 个单位。

调整显示器的黑色电平，以使第一个（负向）元素和参考黑电平难以区分（第二个元素也就基本看不出来）。详见 5.6 节。PLUGE 的负向元素不能用正 $R'G'B'$ 值来表示。

对于标清，现在已撤销的 SMPTE EG 1 对 $\pm 4\%$ PLUGE 进行了标准化，然而 SMPTE RP 219 将高清 PLUGE 标准化为 8 位接口的代码 44、64、84 和 104，即约 -2% 、 $+2\%$ 和 $+4\%$ 。

36.4 使用彩条调整复合解码器

当复合 NTSC 或 PAL 彩条被解码时，从白色、青色和品红中解出的蓝色的数量，以及蓝色条在理想情况下应完全相同。除了从白色中解码的蓝色外，任何色度增益（饱和度）误差会影响到从蓝色解码的信号。色度相位误差会导致从青色和品红色中解出的蓝色有反方向的色调误差。手动调整解码器的色调和色度（或色调和颜色，或相位和饱和度）控制涉及复合 SMPTE 彩条的显示，以及在解码器的输出端禁用红色和绿色通道。调节解码器的色调，可以使从青色和品红解码出的蓝色数量均衡，调整色度可以使从灰色和蓝色中解码出的蓝色数量均衡。SMPTE 彩条的反转彩条部分用于进行比较。图 36-4 示出的彩条图像显

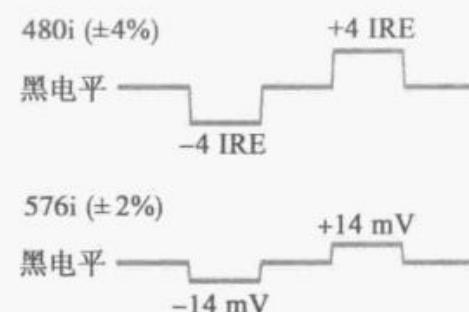


图 36-3 PLUGE 元素

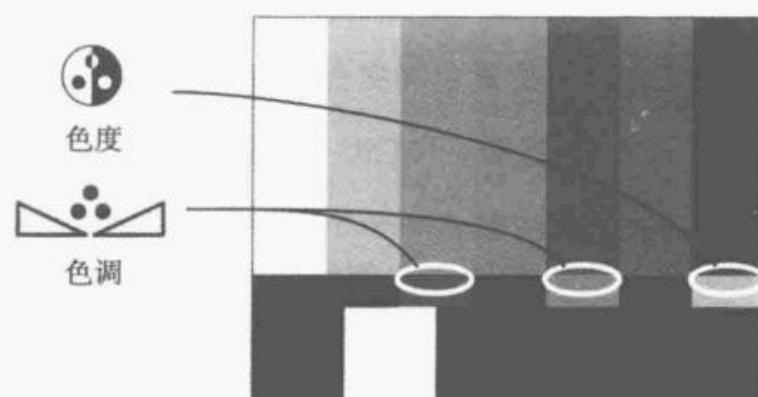


图 36-4 调整色调和色度

示了在调节色调和色度时需要在视觉上比较的横条。

使用彩条测试图像和只显示蓝色的显示器调整复合 NTSC 解码器的色调和色度时，要调整控制端，直到标出的过渡边消失为止。

用这种方式来调节色调和色度控制只对于补偿 NTSC 和 PAL 解码误差有意义。在分量视频中，如 $R'G'B'$ 和 $Y'C_B C_R$ ，不会有录制或传输的损伤使色相旋转或改变色度（饱和度），所以使用上述方案是没有意义的。

36.5 标清彩条中的 $-I$ 、 $+Q$ 和 PLUGE 元素

SMPTE 彩条图案左下象限中包含相位为 $-I$ 和 $+Q$ 的副载波频率成分。这些分量用于确定最初的 NTSC 色度调制方法（约 1953 年）中编码和解码的坐标轴。1970 年左右停止使用由 I 和 Q 确定坐标轴进行编码和解码的方法，改为在 $B'-Y'$ 和 $R'-Y'$ 轴上进行编码和解码，所以这部分信号已不起作用了。传统的 $-I$ 和 $+Q$ 元素包含黑色上的高色度。这些组合对应一个分量为很大的负值的非法 $R'G'B'$ 组合；因此， $-I$ 和 $+Q$ 元素不能在正 $R'G'B'$ 域中表示。

$-I$ 和 $+Q$ 元素分别对应于 $R'G'B'$ 值 $[-0.3824, 0.1088, 0.4427]$ 和 $[0.2483, -0.2589, 0.6817]$ ，以及 10 位 $^{601}Y'C_B C_R$ 值 $[0, 228, -244]$ 和 $[0, 345, 159]$ 。用 $R'G'B'$ 值 $[0, 0.2456, 0.412545]$ 和 $[0.253605, 0, 0.470286]$ 分别产生与 $-I$ 、 $+Q$ 具有相同的色调和饱和度，并且具有最小亮度的合法的 RGB 代码，见 SMPTE RP 219。

$-I$ 元素、 $+Q$ 元素和 PLUGE 的负向元素均为合成信号。这些元素都不代表图像信息，都不能用于评估图像，并且也不能产生于或经过 0 与 1 之间的正 $R'G'B'$ 域（或 0 和 255，或甚至在下余量钳位时 16 到 235 或 64 到 940）。

实际上，不仅仅彩条的 $-I$ 、 $+Q$ 和 PLUGE 元素是合成的，整个信号都是合成产生的！我们称其为彩条信号而不是彩条图像，因为 $-I$ 、 $+Q$ 和 PLUGE 元素不能用非负 RGB 分量来表示。彩条信号表示 $R'G'B'$ 值，就好像这些值是来自经伽马校正的相机并在编码器之前插入一样。彩条的 $R'G'B'$ 值隐含是经过伽马校正的。

36.6 高清彩条

图 36-5 所示的是高清的 SMPTE 彩条测试图案产生的图像。

见 SMPTE RP219《高清、标清兼容的彩条信号》。

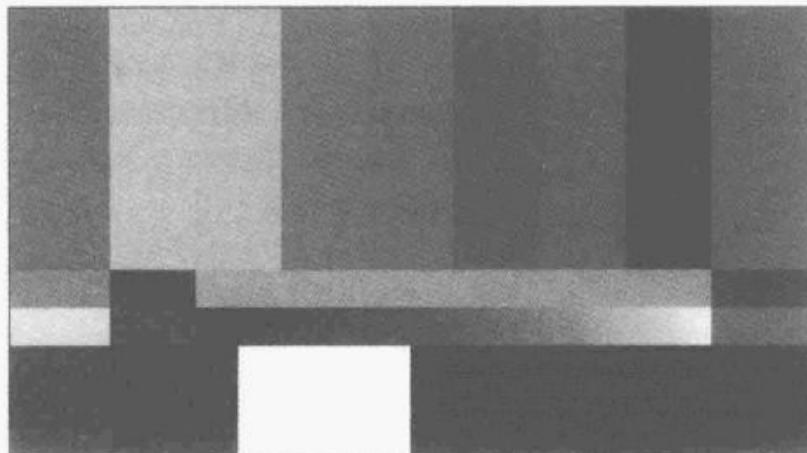
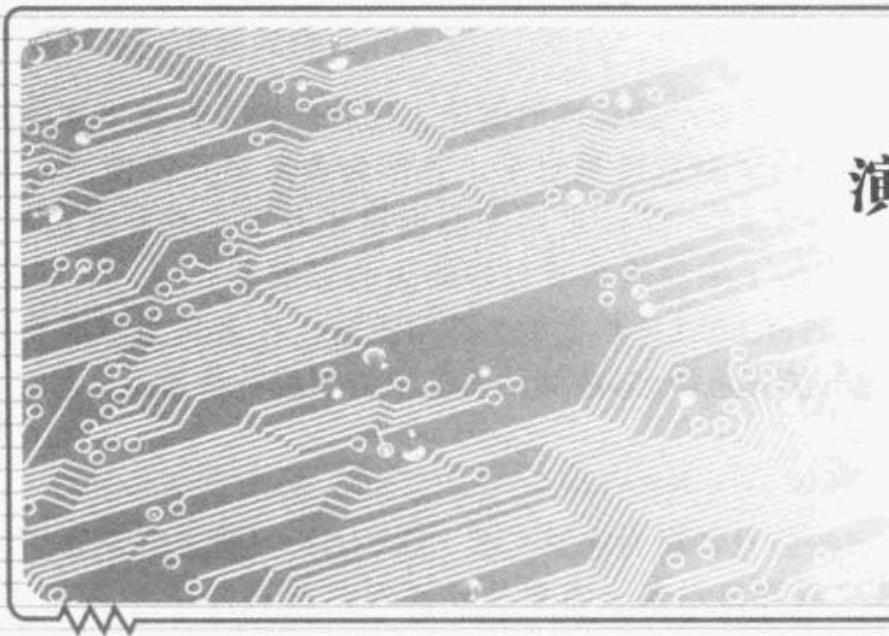


图 36-5 SMPTE RP 219 高清彩条测试信号

第四部分

演播室标准



第37章

参考显示器和观看条件

历史上，视频行业在视频参考显示器标准的制定上比较松懈。在接下来的几年中，随着固定像素显示器（FPD）达到演播室质量水平，并作为参考显示器使用，形势需要改变。在本书写作时，EBU、SMPTE、ITU-R 内的标准组正致力于弥补这些不足。然而，由于当前仍无官方标准，本章接下来的内容就当作缺失的标准来介绍。本章试图以 CRT “广播视频监视器”为范例，将在 2011 年的实践工作写下来。作者认为，液晶显示屏还不是演播室级别。将来需要为容纳宽色域和/或高动态范围控制进行修改。

37.1 简介

参考显示器将 $R'G'B'$ 视频信号转换为根据 CIE 15 的 CIE 三刺激值表征的光信号，就像是通过了加性 RGB 处理（参考显示器只需要从外部观察起来像加性 RGB 设备那样工作，不需要在物理上使用相加混色）。

37.2 信号接口

$R'G'B'$ 视频信号值在此归一化为：参考黑色的值为 0，参考白色的值为 1。低于 0 的信号值钳位到 0。可以容纳最高为参考白色 955/876 的信号值，更大的值将钳位到 955/876（不同的视频标准定义了如何映射到整数。例如，依据 ITU-R BT. 709，10 位数字图像编码包括放大 876 倍并加上 +64 的偏移量；参考黑色用 $R'G'B'$ 信号代码 64 表示，参考白色用 $R'G'B'$ 信号代码 940 表示，峰值白色用代码 1019 表示）。

37.3 参考基色（黑色、白色）

红、绿、蓝、黑、白参考色的三刺激值具有 BT. 709 规定的色度。参考白色信号 $[1, 1, 1]$ 显示照度为 $100\text{cd} \cdot \text{m}^{-2}$ 。参考黑色信号 $[0, 0, 0]$ 相比于参考白色的显示相对照度用 β 表示，其值最好为 0.0003，但不应超过 0.001。参考白色信号和黑色信号应根据 ITU-R BT. 815 来测量。

37.4 参考 EOCF

在参考黑色信号到峰值白色范围内的每一个 $R'G'B'$ 信号分量值（用 V 表示），其对应的基色参考三刺激值（用 T 表示）相对于白色参考信号，按 SMPTE RP 177 计算应为：

$$T = (1 - \beta)(\text{MAX}[0, b + (1 - b) \cdot V])^{2.4} + \beta \quad (b \approx 0.035, \beta \leq 0.0003) \quad (37-1)$$

式中： $1 - \beta$ 为增益； b 为偏移； β 为偏移；2.4 为伽马校正值。

参数 b 取值约为 0.035，由设定黑电平的实际操作来确定（增益因子 $1 - \beta$ 确定了归一化的参考白色，得到的峰值白色照度约为 $122\text{cd}/\text{m}^2$ 。幂函数的指数 2.4 符合 BT. 1886 标准）。

37.5 参考观看条件

参考显示器的环境包括显示表面和其围绕物，在 CIE D₆₅ 色度下其照度应为 $2\pi\text{lx}$ （约 6lx）。参考显示器应当被漫反射系数为 0.5 的近似 Lambertian 中性灰色环绕（这些条件产生 1% 很暗的周边环境）。围绕物应当在水平方向上延展到对应 3 倍画面宽度的角度，在垂直方向上延展到对应 3 倍画面高度的角度。

延伸阅读

CIE 15: 2004, *Colorimetry*, Third Edition (2004) .

ITU-R BT. 709-5, *Parameter values for the HDTV standards for production and international programme exchange* (2002) .

ITU-R BT. 815-1, *Specification of a signal for measurement of the contrast ratio of displays* (1994-07) .

ITU-R BT. 1886, *Reference electro-optical transfer function for flat panel displays used in HDTV studio production* (2011) .

SMPTE RP 177, *Derivation of Basic Television Color Equations* (1993) .

第38章

SDI 和 HD-SDI 接口

本章介绍未压缩和压缩的标清（SD）和高清（HD）的数字接口。表 38-1 和表 38-2 总结了视频信号电平。

表 38-1 几种接口的模拟视频电平总结

接 口	参考黑电平	参考白电平
算术表示的抽象信号值	0	1
IRE 单位表示的抽象信号值	0	100
模拟 NTSC/mV	$53 \frac{4}{7}$	$714 \frac{2}{7}$
模拟 NTSC-J/mV	0	$714 \frac{2}{7}$
模拟 PAL/mV	0	700
零建立电平的模拟 VGA/mV	0	700
7.5% 建立电平的模拟 VGA/mV	$53 \frac{4}{7}$	$714 \frac{2}{7}$

表 38-2 几种接口的数字视频电平总结

接 口	负峰值 非 SDI 黑电平	负峰值 SDI 黑电平	参考黑电平	参考白电平	正峰值 SDI 白电平	正峰值 非 SDI 白电平
8 位计算机技术（例如 sRGB 标准）			0	255		
演播室视频接口 8 位（消费类）	0	1	16	235	254	255
演播室视频接口 10 位（消费类）	0	4	64	940	1019	1023
演播室视频处理 8 位	-16	-15	0	219	254	255
演播室视频处理 10 位	-64	-60	0	876	955	959
数字影院接口 12 位	0	16^{\oplus}	0	3960^{\oplus}	3960	3960

注：①在数字电影中，真正的参考黑色信号无法通过 HD-SDI 接口传输；最小的接口编码支持 $(16/3960)^{2.6}$ 或约 0.000 000 6 的黑色三刺激值，与理想黑色的差异可以忽略。

②在数字电影中，峰值白色编码为 3960；该编码用于亮度 (Y') 通道。另外两个通道 (X' 和 Z') 的峰值分别为 3794 和 3890。

38.1 分量数字标清接口 (BT. 601)

1984 年被采纳的 ITU-R Rec. BT. 601 规定了 480i29.97 和 576i25 标清的抽象编码参数（包括 4:2:2 的色度亚采样）。亮度以 13.5MHz 的速率采样； C_B 和 C_R 色差分量在水平方向上以相对于亮度的 2:1 的因子下采样，即各以 6.75MHz 速率采样。采样值按照 $\{C_B, Y'_0, C_R, Y'_1\}$ 的顺序复用。同步信息和可选的辅助数据也进行复用；27MW/s 的 10 位字经串行化，数据传输速率为 270Mb/s。这种外部接口称为串行数据接口（SDI），使用同轴电缆和 BNC 接头。

ITU-R Rec. BT. 601-5

以 13.5MHz 速率对 480i 系统和 576i 系统采样，都能产生整数个每行的采样点 (S_{TL})，分别为 858 S_{TL} 和 864 S_{TL} 。480i 系统和 576i 系统的每行有效的亮度采样点 (S_{AL}) 都是 720 个。未压缩的 8 位 BT. 601 视频的有效采样点占用大约 20MB/s。

标号 4 : 2 : 2 一开始指一种色度亚采样方案。在 20 世纪 80 年代，4 : 2 : 2 表示一种特定的带有 4 : 2 : 2 色度亚采样的标清分量数字视频接口标准。在 20 世纪 90 年代，4 : 2 : 2 色度亚采样格式用于高清。因此，标号 4 : 2 : 2 变成与图像的尺寸无关。

BT. 601 和 BT. 656 标记在演播室视频中已经不再使用了。然而，桌面视频硬件设计师常使用“Rec. 601”或“BT. 601”标示用分开的线路传输垂直和水平同步信令的并行接口。用“Rec. 656”或“BT. 656”标示用嵌入的 TRS 编码表示垂直和水平同步的并行接口。

图 38-1 示出 480i 分量视频信号一个扫描行的亮度（或 R' 、 G' 、 B' ）波形。时间轴显示在 BT. 601 13.5MHz 采样速率下的采样点记数，采样数除以 13.5 可得到以微秒为单位的时间。幅度单位为 mV（根据 EBU Tech. N10 电平），同时用 8 位 BT. 601 数字接口编码值表示。

图 38-1 显示的是 EBU Tech. N10 模拟电平。然而，这些电平很少在 480i 中使用。在模拟视频中，同步信号比参考黑色更黑，为 -300mV。在 8 位接口中，黑色的编码是 16，白色的编码是 235。720 个有效采样点包含画面信息，在 858 个采样点中，剩下的 138 个采样间隔由水平消隐构成。

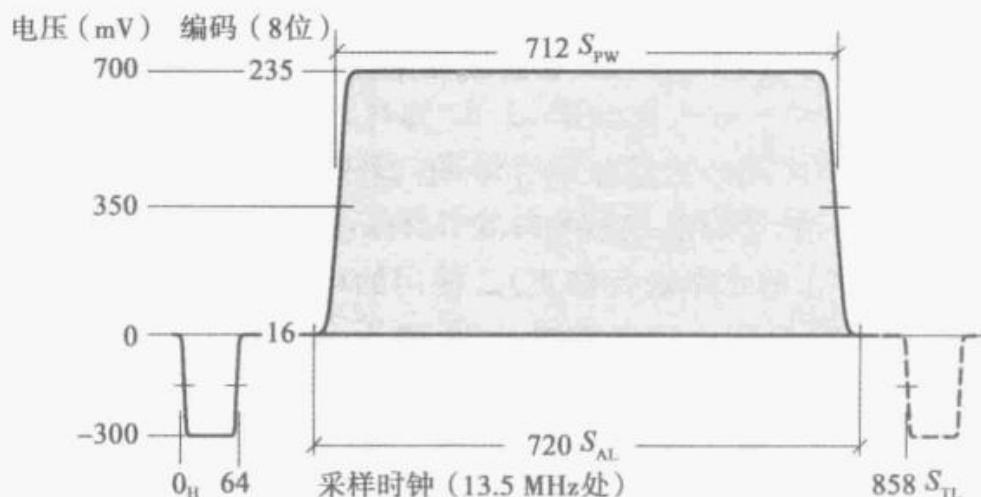


图 38-1 480i29.97 的 4 : 2 : 2 亮度分量扫描行波形

数字视频接口传输被定时基准信号 (TRS) 序列分隔成帧的有效视频，TRS 序列包括有效视频起始 (SAV) 和有效视频结束 (EAV)。在未被有效视频占用的区域可传输辅助数据 (ANC) 和数字化的辅助信号。

图 38-2 所示为第 8 章的光栅图加上 EAV、SAV，以及 HANC 和 VANC 区域，更详细的内容见 38.4 节。

BT. 656 分量数字接口使用 EAV 指示每个水平消隐间隔的开始，用 SAV 指示有效视频的开始。在 EAV 和 SAV 之间，可携带辅助数据 (HANC)。在非画面行，SAV 和 EAV 之间的区域可携带辅助数据 (VANC)。数字化的辅助

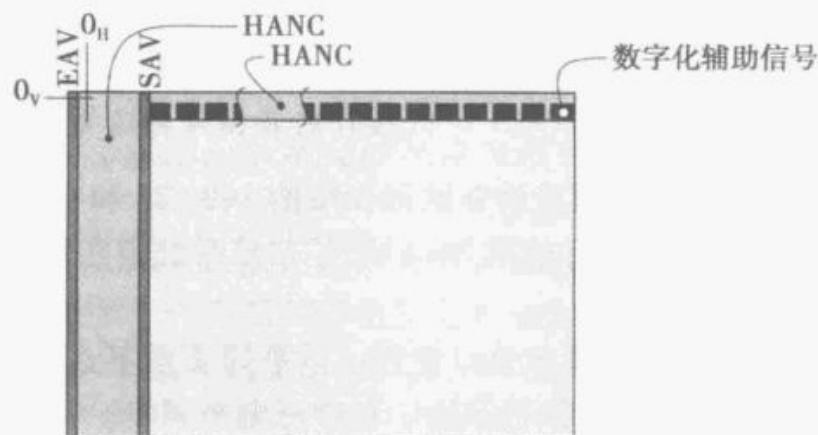


图 38-2 BT. 656 分量数字接口

信号可以在不传输 VANC 或模拟同步的行中携带。

38.2 串行数字接口

串行数字接口 (SDI) 指由 SMPTE 标准化的一组接口。BT. 601 或者 $4f_{sc}$ 数据流先串行化，然后用加扰技术处理。SMPTE ST 259 (《10 位 4:2:2 分量和 $4f_{sc}$ 复合数字信号——串行数字接口》) 标准化了几种接口，用字母 A 到 D 标识，如下所示：

- 复合 $4f_{sc}$ NTSC 视频，数据传输速率约 143Mb/s。
- 复合 $4f_{sc}$ PAL 视频，数据传输速率约 177Mb/s。
- BT. 601 4:2:2 分量视频，数据传输速率为 270Mb/s (该接口在 BT. 656 中标准化)。
- 以 18MHz 采样的 BT. 601 4:2:2 分量视频，实现 16:9 的幅型比，数据传输速率为 360Mb/s。

除了 C 方案，其他的都已过时。

SDI 为通过同轴电缆的电气传输和通过光纤的传输进行了标准化。SDI 电气接口使用 ECL 电平， 75Ω 阻抗，BNC 连接器和同轴电缆。在 SMPTE 标准和 BT. 656 中指定了电气和机械参数，见 38.9 节。SMPTE 297M 的数字标清光纤接口直接适应 BT. 656 的串行版本。

38.3 分量数字 HD-SDI

高清系统的基本编码参数在 BT. 709 中被标准化。各种扫描系统详见表 15-2 引用的若干 SMPTE 标准。

分量标清、复合 $4f_{sc}$ NTSC 和复合 $4f_{sc}$ PAL 制式都有不同的采样率和不同的串行接口数据传输速率。对于高清，采用一致的 74.25MHz 采样速率（在需要兼容 59.94Hz 帧频的应用中，用 1000/1001 的比率进行修正）。使用的串行接口的数据传输速率为采样率的 20 倍。相同标准的不同变体适应主流的 1080i30/1080p24/720p60 扫描、1080p30 和已被淘汰的 1035i30 系统。整数画面频率 24、30 和 60 可用分数 1000/1001 修正，得到 23.976Hz、29.97Hz 和 59.94Hz 的频率。

为了适应高清，270Mb/s 的 SDI 接口将数据传输速率乘以 5.5 倍，得到 1.485Gb/s 的固定数据传输速率。23.976Hz、29.97Hz、59.94Hz 接口的采样率和串行数据传输速率如下所示。该接口标准化为 $Y'C_B C_R$ ，4:2:2 下采样。双链路 HD-SDI 可以用来传输 $R' G' B' A$ ，4:4:4:4。

23.976Hz、29.97Hz 和 59.94Hz 帧频和采样速率的关系是： $\frac{74.25}{1.001} \approx 74.176\text{Mpx/s}$

对应的 HD-SDI 串行接口的数据传输速率是： $\frac{1.485}{1.001} \approx 1.483\text{Gb/s}$

HD-SDI 可容纳在欧洲使用的 1080i25 和 1080p25 的变体。这是通过将 1920×1080 像素图像阵列置于速率为 25Hz 的扫描系统中实现。 S_{TL} 修改自 30Hz 标准，形成了 1125/25 光栅。

标准高清模拟接口使用三电平同步而不是模拟标清使用的两电平同步。图 38-3 示出 1080i30 高清扫描行波形，包含三电平同步。

图 38-3 显示了模拟三电平同步，幅度为 $\pm 300\text{mV}$ 。在 8 位接口中，黑色由编码 16 表示，白色由编码 235 表示。显示的 1920 个有效采样点包含画面信息，总共 2200 个采样点。

中剩余的采样间隔由水平消隐构成。

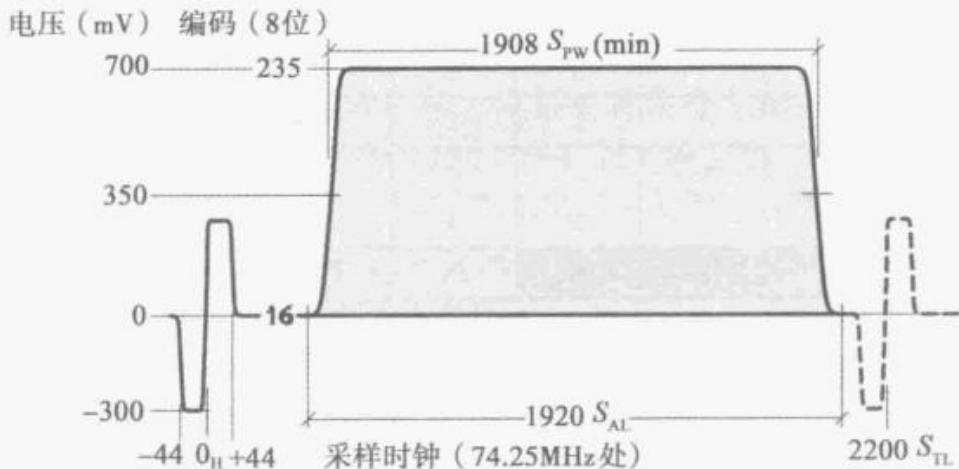


图 38-3 1080i30 高清亮度分量的扫描行波形

SMPTE ST 292 (《Bit-Serial Digital Interface for High-Definition Television Systems》) 标准化了 HD-SDI 接口，也指定了数字高清的光纤接口。

38.4 SDI 和 HD-SDI 同步、TRS 和辅助数据

伴随着图像信息，流式接口需要传输与每帧起始和每行起始相关的时刻或数字采样点的信息。在数字视频中，这些信息通过定时基准信号 (TRS) 传输，本章将详细解释。

SDI 有传输辅助数据 (ANC) 的容量。串行数据传输接口 (SDTI) 类似于 SDI，但不含未压缩的有效视频。相反，链路的全部数据容量都用于携带 ANC 数据包，压缩的数字视频可以通过这些数据包传输。SDTI 将在 38.12 节介绍。

IEEE 1394/DV 接口和 DVB-ASI 接口都没有辅助数据，也不使用 TRS。

标准串行接口传输 10 位采样值；发送器必须在接口呈现所有 10 位数据（即使两个最低有效位为零）。

TRS 和 ANC 序列的开头是 10 位码字 0 和 $(3FF)_h$ 。从 SMPTE RP 125 和 EBU Tech. 3246 这样的老式并行接口开始，接收机在识别 TRS 和 ANC 时必须忽略两个最低位。除了用于界定 TRS 和 ANC，码字 0、1、2、3 和 $(3FC)_h$ 、 $(3FD)_h$ 、 $(3FE)_h$ 、 $(3FF)_h$ 在数字视频数据中是禁用的。

38.5 4:2:2 SD-SDI 中的 TRS

38.1 节介绍了 4:2:2 采样值按 $\{C_B, Y'_0, C_R, Y'_1\}$ 序列复用到 SDI 上。BT. 601 定义了抽象信号编码参数，BT. 656 定义了接口。

有效亮度采样值从 0 到 $S_{AL}-1$ 编码，有效色度采样值从 0 到 $(S_{AL}/2)-1$ 编码。接口在每个亮度采样时钟传输两个字：偶数字传输色度采样值，奇数字传输亮度采样值。采样结构与 0_H 对齐，如果将模拟同步数字化，一个特别的数字化的亮度样值将精确地反映同步值的 50%。

在 4:2:2 视频中，4 字的 TRS 序列在有效视频之前，说明有效视频起始 (DAV)。 C_B 的首个采样值跟在 SAV 之后。紧随这一行最后一个有效采样值的是另一个 4 字 TRS 序列，有效视频结束 (EAV)。TRS 序列包含了一个全 1 的字 $(3FF)_h$ 、两个全 0 的字，最后是一个含有标志位 F (场)、V (垂直)、H (水平)， P_3 、 P_2 、 P_1 和 P_0 (校验) 的字。 $H=0$ 表示 SAV， $H=1$ 表示 EAV。

V 和 H 位在相应的消隐间隔里声明。F 位表示场而不是帧。

表 38-3 显示了 TRS 元素。

表 38-3 定时基准序列 (TRS)

字	值	MSB 9	8	7	6	5	4	3	2	1	LSB 0
0	(3FF) _h	1	1	1	1	1	1	1	1	1	1
1	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0
3		1	F	V	H	P_3	P_2	P_1	P_0	0	0

在相关联的行开始之前的 EAV 中 F 和 V 位会改变状态，与其叫作有效行结束，不如称为水平间隔起始。在隔行扫描系统中，F 在第二场中为 1。在 480i 系统中，F 在第 4 行和第 266 行变化，在其他扫描系统中，包括 576i 和高清，F 在第 1 行改变状态。在逐行系统中，F 总是 0（除了在 483p59.94 中，F 对帧校验编码）。

在每一个由相关扫描标准定义为含有有效（画面）视频的行中，垂直消隐（V）位是 0；在别处，即垂直间隔中，为 1。

在 BT.601-4 (1994) 和 SMPTE RP125-1992 中，在 480i 的系统中，V=0 的 SAV 可能在第 1 个有效（画面）行之前出现，早在第 10 行和第 273 行。为了与老式设备兼容，不要依赖 V 从 1 到 0 的变化。

F、V 和 H 位由奇偶校验位 P_3 、 P_2 、 P_1 和 P_0 来保护，奇偶校验位是对 F、V 和 H 中的两个或者三个求“异或”而得，如表 38-4 所示。

表 38-4 SAV 和 EAV 的保护位

值	F	V	H	$P_3 = V \oplus H$	$P_2 = F \oplus H$	$P_1 = F \oplus V$	$P_0 = V \oplus H$		
(200) _h	1	0	0	0	0	0	0	0	0
(274) _h	1	0	0	1	1	1	0	1	0
(2AC) _h	1	0	1	0	1	0	1	1	0
(2DB) _h	1	0	1	1	0	1	1	0	0
(31C) _h	1	1	0	0	0	1	1	1	0
(368) _h	1	1	0	1	1	0	1	0	0
(380) _h	1	1	1	0	1	1	0	0	0
(3C4) _h	1	1	1	1	0	0	0	1	0

各种 SMPTE 标准对有效区域外的字的编号是不一致的。EAV 通过 F 和 V 位的状态改变起到标示数字行起始的作用，所以作者在 EAV 时从 0 开始对字编号。在这个方案中，SAV 开始于字 $S_{TL} - S_{AL} - 4$ 。将 EAV 编号为字 0 的另一个原因是提出的 HD-SDTI SMPTE 标准 (SMPTE ST 348) 设计了一种提前 SAV 编码时序的方案。字和采样点都用严格的计数进行编号，无论是字还是采样点编号都不出现在接口上。

接口上从 EAV 到 SAV 的水平消隐间隔可以包含辅助数据 (HANC)。在每一个有效（画面）行，从 SAV 到 EAV 的间隔包含有效视频。在有效画面行之外的 SAV 和 EAV 之间的间隔可以用于传输辅助数据 (VANC)。如果在有效画面外的一行没有携带 VANC，并且这行与模拟同步元素没有关联，从 SAV 到 EAV 的间隔可携带像有效视频那样编码的数字化的辅助信号。不用于 EAV、SAV、有效视频、数字辅助信号或者辅助数据 (ANC) 的间隔将以在有效画面中代表消隐的交替的代码 {色度 (200)_h，亮度 (40)_h} 来填充。

38.6 HD-SDI 中的 TRS

HD-SDI 与 4 : 2 : 2 SD SDI 类似，只是单个链路携带两个逻辑流，一个是色度，另一个是亮度。每一个流都含有 TRS 序列，每个流都可携带独立的 ANC 包。两个流是字复用的，复用流经过串行化和加扰。表 38-5 所示的 4 个字附在每个 EAV 之后。每个 bit9 是 bit8 的补数。字 4 和字 5 传输行号 (LN0, LN1)。字 6 和字 7 提供流中有效视频的 CRC 保护。每一个流都有一个实现 $x^{18} + x^5 + x^4 + 1$ 特征函数的 CRC 发生器。在 SAV 之后，每一个发生器立即复位为 0，并将字累加到最大为 LN1。

CRC 覆盖了从第一个有效采样值一直到行号。一个 HD-SDI 接口传输两个流，每个都包含 EAV+LN+CRC 和 SAV 序列；一个流携带色度对齐字，另一个流携带亮度对齐字。

表 38-5 HD-SDI 的行号和 CRC (这个数据包记作 EAV+LN+CRC)

字	值	MSB 9	8	7	6	5	4	3	2	1	LSB 0
4	LN0	\bar{L}_6	L_6	L_5	L_4	L_3	L_2	L_1	L_0	0	0
5	LN1	1	0	0	0	L_{10}	L_9	L_8	L_7	0	0
6	CR0	$\overline{\text{CRC}_8}$	CRC_8	CRC_7	CRC_6	CRC_5	CRC_4	CRC_3	CRC_2	CRC_1	CRC_0
7	CR1	$\overline{\text{CRC}_{17}}$	CRC_{17}	CRC_{16}	CRC_{15}	CRC_{14}	CRC_{13}	CRC_{12}	CRC_{11}	CRC_{10}	CRC_9

38.7 模拟同步和数字/模拟时序关系

在模拟视频中，同步用“比黑色更黑”的视频电平来传输。行同步通过给每一个行扫描关联一个记作 0_H 的（水平）行同步数据来实现， 0_H 定义在同步先导沿（下降沿）的中点。场和帧同步通过给每一场关联一个记作 0_V 的垂直同步数据来实现。

同步分离恢复出和模拟视频信号相关的重要时间点。同步锁定重建一个采样时钟。

数字和模拟域的联系是建立在 0_H 相对于某些 TRS 元素的位置的基础上的。表 38-6 总结了几种数字表示分量 4 : 2 : 2 视频的标准。最右边一列给出了在 EAV 的第一个字和 0_H 采样值（如果其被数字化）之间的亮度采样间隔的个数。

表 38-6 数字与模拟的时序关系

系 统	AR	扫 描	标 准	S_{π}	S_{AL}	EAV 到 0_H
483i29.97		525/59.94/2 : 1	SMPTE 125M, BT. 601	858	720	12
576i25		625/50/2 : 1	EBU 3246, BT. 601	864	720	16
483i29.97	16 : 9	525/59.94/2 : 1	SMPTE 267M, BT. 601	1144	960	16
483p59.94	16 : 9	525/59.94/1 : 1	SMPTE 293M	1144	960	16
576i25	16 : 9	625/50/2 : 1	EBU 3246, BT. 601	1152	960	21
720p60	16 : 9	750/60/1 : 1	SMPTE ST 296	1650	1280	110
1080i30	16 : 9	1125/60/2 : 1	SMPTE ST 274	2200	1920	88
1080p30	16 : 9	1125/30/1 : 1	SMPTE ST 274	2200	1920	88
1080p25	16 : 9	1125/25/1 : 1	SMPTE ST 274	2640	1920	192
1080p24	16 : 9	1125/24/1 : 1	SMPTE ST 274	2750	1920	192

38.8 辅助数据

在 4 : 2 : 2 标清和高清中，辅助数据允许在一个既不包含有效画面也不包含数字化的

辅助数据的行中紧接着任何 EAV (HANC) 或是 SAV (VANC) 后面传输 (在 576i 中, 辅助数据限制在 20~22 行和 333~335 行中)。一个辅助包的开头是一个由 {0, (3FF)_h, (3FF)_h} 三字序列构成的辅助数据标志 (ADF)。

要确定一行是否确实包含数字化辅助信号, 要检查每个色度/亮度对是否为 {(200)_h, (40)_h}: 如果任何一对不等于这组值, 当前的就为数字化辅助信号。

一个 ANC 包不能干扰有效视频或任何 TRS、SAV、EAV。允许有多个 ANC 包, 只要它们是连续的。特定的 ANC 区域为特定用途保留, 参考 SMPTE ST 291。

一个辅助包包括 3 个字 (4 : 2 : 2) 或 1 个字 ($4f_{SC}$) ADF, 后面跟着以下元素:

- 一个字的数据 ID (DID)。
- 一个字的数据块序号 (DBN) 或第二级 DID (SDID)。
- 一个字的数据计数 (DC), 0~255。
- 0~255 个用户字 (UDW)。
- 一个字的校验和 (CS)。

每个包头字——DID、DBN/SDID 和 DC, 是一个 8 位值。每个包头字的 bit8 是奇偶校验位, 当 bit7~bit0 有奇数个 1 时为 1, 当有偶数个 1 时为 0。bit9 是 bit8 的补数 (从而避免了最高 8 位为全 0 或全 1 的码字, 不至于与引导 TRS 和 ANC 序列的 0_H 和 (3FF)_h 码冲突)。

两种类型的 ANC 包在 DID 字的 bit7 上有区别。如果 DID₇ 被声明 (为 1), 这个包是类型 1; 数据块编号 (DBN) 跟在 DID 后面。类型 1 的包可用的 DID 码有 128 种。

DBN 值表示连续性: 如果为 0, 表示未被使用; 反之, 用于对每个 DID 中的包从 1 到 255 计数, 模为 255。

如果 DID₇ 为 0, 包为类型 2: DID 后面跟着二级数据 ID (SDID), 给类型 2 的包提供了 127×255 (即 32 385) 个 ID 编码。

三个 DID 值——(004)_h、(008)_h 和 (00C)_h 指示用 8 位数据编码的类型 2 的 ANC 包。(001)_h~(00F)_h 范围内的其他值是禁用的。DID (80)_h 标记了待删除的包。DID (84)_h 标记了在 VANC 或 HANC 区域中最后一个 ANC 包。

数据计数 (DC) 字包含了一个从 0 到 255 的值 (由两个奇偶校验位保护), 表示了用户数据区中字的个数。DC 字用到了接口的全部 10 位。即使表示的是 8 位的 DID, SMPTE 标准也默认 DC 字的两个最低有效位是有意义的 (如果最低两位无意义, 用户数据字的计数便不能唯一确定)。

校验和 (CS) 字为辅助包的内容提供了完整性校验。从 DID 到 UDW 的最后一个字, 每个字的最高有效位 (MSB) 将被屏蔽掉 (为 0), 这些值以模 512 求和。9 位的和用 CS 的 bit8 到 bit0 传输; bit9 是 bit8 的补数。

38.9 SDI 编码

在串行接口中, 接收端从编码的比特流中恢复时钟是必需的。因此编码的比特流必须在编码比特率上有显著的功率。同轴电缆衰减高频信息, 需要用均衡器来克服这种衰减。因为均衡器有高频交流电路, 编码的比特流应该在很低频率上含有很小的功率: 编码应该是无直流的。为了实现经济的均衡器, 用于正确恢复信号的频率范围应尽可能小。大约 2 : 1 的最高和最低的频率分量比率——编码信号包含在 1 倍频程带宽内——是可取的。这些考虑要求有高时钟速率。不过, 显然采用低时钟速率成本最低。时钟速率的选择就是在这些需求中取折中。

SDI 使用加扰编码，数据流经串行化，然后通过一组带有“异或”抽头，实现 $x^9 + x^4 + 1$ 特征函数的移位寄存器。

使用一个加扰器的扰码技术为众人熟知，但是 SDI 和 HD-SDI 还用了一个二级加扰器，其特征函数为 $x+1$ 。级联的两级加扰器实现了比传统的单级加扰器更好的效果。扰码技术是自同步的，不需要初始化。

接口的数据传输速率是字速率乘以每个字的位数得到的。标准的做法是将 10 位数据串行化，当编码 8 位视频时，两个最低有效位置为 0。没有规定能够避免在串行化和加扰之后出现一长串连 0 或连 1 的串行位序列，但一长串连 0 或连 1 无法提供使接收端能够恢复时钟的信号转换！在实际中，这种病态序列是很少见的。

SDI 被标准化为通过同轴电缆进行 ECL 电平的电气传输，使用如图 38-4 所示的 BNC 接头。可实现的传输距离为 200~400m。SDI 也被标准化为通过光纤传输。光纤数字 SD 接口是 SDI 的直接适配版本。



图 38-4 BNC 接头

38.10 HD-SDI 编码

270Mb/s 的 SDI 接口将数据传输速率放大 5.5 倍，以用于高清传输，得到 1.485Gb/s 的数据传输速率（或者在 24.976、29.97、59.94Hz 的系统中为 1.485/1.001Gb/s），这两种数据传输速率都称为“1.5Gb/s”。SMPTE ST 292 制定了 HD-SDI 标准。HD-SDI 仿照 SD SDI，但是有两点重要的改变：

- 色度和亮度在分开的两个数据流中编码，各自都有自己的 TRS 序列。该数据流被复用，然后被加扰。
- 编码的行号和 CRC 附加到 TRS 序列的 EAV 部分（叫作 EAV+LN+CRC）。

在 1080i 和 1080p 标准文档中，采样值相对于 0_H 来编号（不像标清标准的文档，采样值相对于每行第 0 个有效采样来编号）。

后续的发展将 1.5Gb/s 接口的数据传输速率翻倍至约 3Gb/s（有时称为 3G-SDI）。这个速率足以传输 4：4：4：4（即没有色度下采样，有 α 分量）的 1080i30 或 4：2：2 的 1080p60。

38.11 压缩视频接口

由于压缩系统的多样性以及压缩接口需要信号处理和显示设备具备解压缩能力，压缩数字视频接口在演播室中并不实用。压缩的 4：2：2 数字视频演播室设备通常通过非压缩的 SDI 接口互联。

压缩接口可用于向非线性编辑系统传输视频，以及在共用相同的压缩系统的录像机之间进行复制。压缩视频可以直接使用稍后会介绍的串行数据传输接口（SDTI）相连。在网络或传输应用中（但不用于内容制作），DVB ASI 接口广泛用于承载 MPEG-2 传送流。IEEE 1394/DV 接口，有时候也称火线或 i.LINK，广泛用于消费电子舞台，并开始部署在广播应用中。

38.12 SDTI

SMPTE 标准化了一种 SDI 的衍生品，即串行数据传输接口（SDTI），用于在非压缩有效视频的位置传输任意的数据包。SDTI 可用于传输 DV25、DV50、DV100 和索尼 MPEG IMX 压缩数据流。尽管 DV 位流已经标准化，不同生产制造商已经选择了互不兼容的技术将他们的压缩视频数据封装进 SDTI 流。这使得 SDTI 仅对于同一家厂商设备的

互联有用。

38.13 切换和混合

在视频源之间的切换和编辑——“剪切”——是在垂直间隔期间完成的，因而生成视频中的每一帧依然完好，没有任何切换的变化。当用硬件切换器在两个信号间切换时，若要使输出信号在切换瞬间连续，输入信号必须是同步的——两个信号的 0_V时刻必须在时间上精确匹配。为了避免切换时干扰垂直同步元素，切换应该稍晚于 0_V，见 SMPTE RP 168。

38.14 数字设备的时序

现代数字视频设备在每一个输入端有一个用作 FIFO 队列的缓冲区。在每个输入端的缓冲区可容纳多达好几行时间的输入时序（相对于参考视频）超前量。数字设备的定时涉及将每个信号源提前，这样来自所有信源的信号都将及时到达设备主切换器的输入端。这种定时不需要准确，能保证不出现缓冲区上溢或下溢即可。当路由切换器在 SDI 流之间切换时，几十个采样点的时序误差是可以容忍的，下游设备可在切换后的一两行内恢复时序。

当演播室需要容纳异步视频输入时——输入的帧频在容许范围内，但其相位不能用作同步基准，例如卫星信号输入——帧存储同步器将被使用。该设备包含用作视频 FIFO 缓冲区的帧存储器。一个有着任意时序的输入信号按照自己的同步元素的时序写入存储器中。同步器接受一个参考视频信号，存储器以锁定到该参考视频的同步元素的速率读出（制定了调节系统相位——相对于参考视频的输出信号时序的规定）。因而为了匹配本地参考，一个异步信号将延迟多达一帧的时间，有时还会稍长一点，然后信号可像本地信号源一样使用。

演播室视频设备常常包括帧存储器，并表现出一场、一帧或更久的延迟。这类设备的低级时序通过引入时序提前量来实现，使得 0_V出现在正确的时刻。然而，即使视频内容相对于 0_V有正确的定时，也可能是晚了一帧，或在大型设备中晚了数帧。必须注意将音频延迟差不多的时间间隔，以避免唇同步问题。

38.15 ASI

在广播设备内部，MPEG-2 传送流可以串行化到专用的异步串行接口（ASI）上。广播用的串行化 ASI 流有大约 20Mb/s 的净荷数据传输速率。然而，为了能使用 SDI 分发基础架构，ASI 接口的数据传输速率是 270Mb/s。ASI 接口使用 BNC 接头和同轴电缆。ASI 是极性敏感的（不像 SDI），虽然现代的 ASI 接收器通常会检测和纠正极性反转。

虽然使用了 SDI 物理层，串行 ASI 流没有 TRS 码，而且接口没有使用 SDI 加扰。相反，信道数据是按借鉴光纤信道标准的 8b/10b 方案来编码的（因此 ASI 接口数据传输速率最高大约为 216Mb/s）。一个 8b/10b 位流从不会出现超过 4 个的连 0 或连 1，所以时钟恢复比较简单。一个 8b/10b 编码器将介质上的低频（“直流”）分量最小化。

由于 ASI 净荷传输速率通常远远低于信道容量，填充码被插入以占用空闲时间。填充码——光纤信道的逗号编码，记作 K28.5——要么在最早的时机插入（按字节，即字节间隔模式），要么在当前数据包完成时插入（按数据包，即突发模式）。MPEG 包最少用两个逗号码分隔。

用 IP 协议通过以太网来承载传送流的做法变得越来越普遍。

38.16 数字接口总结

表 38-7 总结了标清和高清数字接口标准。

表 38-7 标清和高清数字接口标准

ITU-R Rec. BT. 656, “4:2:2 级的 525 行和 625 行电视系统数字分量视频信号接口, ITU-R BT. 601 建议”
SMPTE 125M, “4:2:2 分量视频信号——位并行数字接口”
SMPTE 259M, “10 位 4:2:2 分量和 $4f_{sc}$ 复合数字信号——串行数字接口”
SMPTE 267M, “位并行数字接口——4:2:2 分量视频信号 16×9 幅型比”
SMPTE ST 292, “高清晰度电视系统的位串行数字接口”
SMPTE ST 297, “ANSI/SMPTE 259M 信号的串行数字光纤传输系统”

第39章

480i 分量视频

本章详细介绍 480i29.97 (525/59.94/2 : 1) 视频的扫描、时序、同步结构和图像结构。本章介绍的扫描和时序信息适用于 480i 视频所有的变体，包括模拟视频和数字视频。同步信息涉及模拟、复合模拟和复合数字系统。

39.1 帧频

480i 视频表示的是以每秒 $30/1.001 (\approx 29.97)$ 帧的恒定速率采样的静止或运动的二维图像。对于演播室视频，帧频的公差一般为 $\pm 10\text{ppm}$ 。实际中，误差适用于一个高频率的主时钟，但为了计算和标准的书写，引用帧频的公差比较方便。

39.2 隔行扫描

1 帧是通过一组 525 条相等时长的水平光栅线传输的，按 2 : 1 隔行均匀地从上到下，从左到右进行扫描，从而形成了第一场和第二场；第二场的扫描行相对于第一场的扫描行在垂直方向上有半个垂直采样间距的偏移，在时域上有半帧时间的滞后。用 MPEG-2 的术语表示，第一场就是底场。

把场称为偶场和奇场容易混淆。应当称为第一场和第二场。

在一整帧中，各行从 1 开始连续编号。

表 39-1 显示了 480i 视频帧的垂直结构，并标示出行号的分配和各行的内容。

表 39-1 480i 行分配

行号，第一场 ($F=0$)	行号，第二场 ($F=1$)	V	内容 (左半部分)	内容 (右半部分)
	266 [3]		EQ	BR
4			BR	BR
	267 [3]		BR	BR
5			BR	BR
	268 [5]		BR	BR
6			BR	BR
	269 [6]		BR	EQ
7			EQ	EQ
	270 [7]		EQ	EQ
8			EQ	EQ
	271 [8]		EQ	EQ
9			EQ	EQ
	272 [9]		EQ	无

(续)

行号, 第一场 ($F=0$)	行号, 第二场 ($F=1$)	V	内容 (左半部分)	内容 (右半部分)
10~19	273~282 [10~19]	$V=0^{\dagger}$ (478 行)	垂直间隔视频 (10 行)	
			垂直间隔视频 (10 行)	
20			垂直间隔视频	
			垂直间隔视频	
21			CC	
			CC	
22			图像	
			图像	
23~261			图像 (239 行)	◊
			图像 (239 行)	
262	286~524 [23~261]		图像	
			图像	
263	525 [262]		图像	EQ
			EQ	EQ
‡264 [1]			EQ	
EQ			‡2	EQ
‡264 [2]			EQ	EQ
EQ			‡3	EQ

注: EQ 均衡脉冲。

BR 宽脉冲。

CC 隐匿字幕。

[n] 相对于第二场起始的行编号 (弃用)。

§ 在 RP 125-1992 和 BT. 601-4 (1994) 的 480i 版本中 $V=0$; 在以后的标准中, 这些行的 $V=1$ 。

◊ 右侧的粗竖条表示按照 SMPTE RP 20 标准在 480i 或 480p MPEG-2 系统中传输的行 (画面的垂直中心位于 404 行和 142 行中间)。不幸的是, 480i DV 系统对从此处向上的一个单像素的图像行进行了数字化。

‡ 在模拟电视系统中, 第 1 行到第 3 行都被认为是第一场的一部分, 第 264 行和第 265 行被认为是第二场的一部分。

在老式设备中, 图像可能早在第 20 行或第 283 行开始。然而, 在第 21 行和第 284 行上的视频会被 NTSC 传输的第 21 行隐匿字幕数据取代, 所以在演播室中提供多于 483 个图像行是没有意义的。随着 480i DV 和 MPEG-2 系统的广泛使用, 作者认为提供多于 480 行并无意义; 然而, 在模拟 NTSC 制式中有 483 行被广播。

关于隐匿字幕, 见 ANSI/EIA/CEA-608-B。

从第 10 行到第 21 行, 以及从第 273 行到第 284 行可能包含与图像有关或无关的辅助 (“垂直间隔”) 信号。如果使用了垂直间隔时间码 (VITC), VITC 应位于第 14 (277) 行; 第二个 (冗余副本) 可置于第 16 行。如果第 14 行不可用, 建议用第 18 (281) 行。见 33.5 节。

关于 VITC 行分配的细节见 SMPTE RP 164。

39.3 行同步

水平事件都是参照标记为 0_H 的时刻。在模拟域, 0_H 定义为每个行同步脉冲的先导 (负跳变) 沿的 50% 点。在分量数字接口中, 同步与数字信息之间的对应关系由接口上传输的定时基准信号 (TRS) 来确定 (见 38.4 节)。

在模拟接口中，每一行随着同步脉冲的负跳变沿开始于 0_H 。除了将要介绍的垂直同步行之外，每一行都起始于一个正常同步脉冲。起始于正常同步脉冲的每一行可能含有视频信息。除了被同步脉冲占用的间隔外，每一个起始于不是正常同步的同步脉冲的行保持消隐电平。

39.4 场/帧同步

为了定义垂直同步，帧分成半行周期的间隔。每个半行要么没有包含同步信息，要么开始于同步脉冲出现。同步脉冲的持续时间为以下三种之中的一种，有 $\pm 0.100\mu s$ 的公差：

- 持续时间为 $4.7\mu s$ 的正常同步脉冲。
- 持续时间为正常同步脉冲一半的均衡脉冲。
- 持续时间为行时间的一半减去正常同步脉冲时间的宽脉冲。

在美国的 EIA 和 FCC 标准把均衡脉冲的持续时间取整为 2 位数字的 $2.3\mu s$ ，比 $2.35\mu s$ 的理论值略小。

场中每一组 525 个半行都开始于如下的垂直同步序列：

- 6 个前均衡脉冲。
- 6 个宽脉冲。
- 6 个后均衡脉冲。

第 263 行起始于一个正常同步脉冲并在行中间有一个均衡脉冲。第 272 行起始于均衡脉冲并保持消隐，中间没有同步脉冲。

垂直事件都参照标记为 0_V 的时刻。在模拟域， 0_V 定义为与 0_H 同时的第一个均衡脉冲。行号 1 用 0_V 表明；在一帧中行号是按隔行的时间顺序而不是空间顺序来计数的。480i 系统独有地在第 1 个均衡脉冲确定 0_V 和第 1 行，在 576i 和高清中， 0_V 和第 1 行标记在第 1 个宽脉冲上。

历史上，在模拟域，每场的 0_V 定义为第一个均衡脉冲的 50% 点。第一场中的行从 1—263 编号，第二场中的行从 1 到 262 编号。在数字域，第一场包含 262 行，第二场包含 263 行。

图 39-1 显示了同步结构的细节，该波形图是表 39-1 的模拟表示。该图显示了第一场和第二场的波形，给出了垂直同步间隔的细节。CC 表明该行可能包含 EIA608 (第 21 行) 隐匿字幕数据。

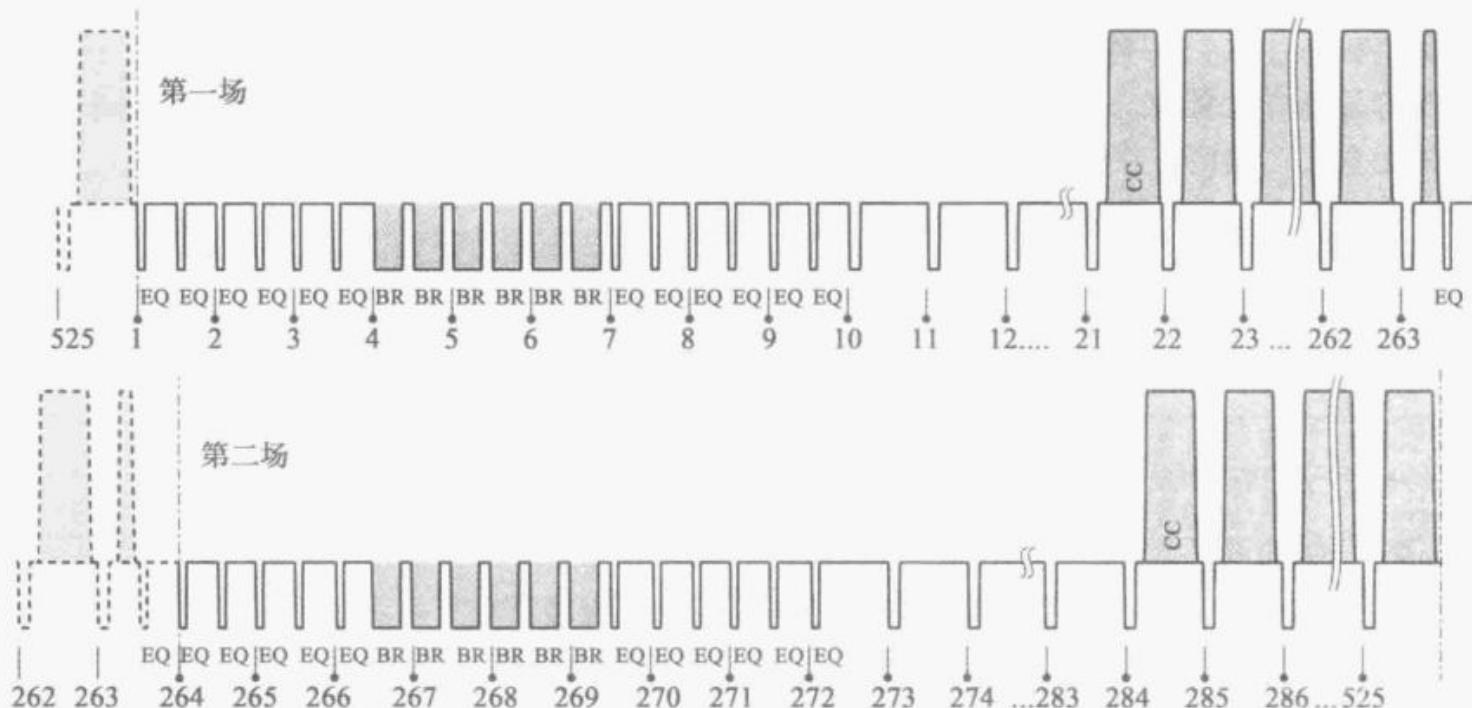


图 39-1 480i 垂直光栅

当同步信号用模拟或数字化的 $4f_{SC}$ 形式表示时，会有一个上升时间（从 10% 到 90%）为 (140 ± 20) ns 的升余弦转换沿，转换沿的中点与理想的同步信号位置相同。

为了避免破坏同步或者图像，视频源间的切换在垂直间隔中进行。切换发生在每场第一个标准行的 0_H 的 (30 ± 5) μs 后。在 480i 系统中，切换在第 10 行当中进行（如果第二场占主导，切换将在第 273 行当中进行）。

见 SMPTE RP 168 “同步视频切换中垂直间隔切换点的定义”

39.5 $R'G'B'$ EOCF 和三基色

图像信息参照线性光基本的红、绿、蓝 (RGB) 三刺激值，用范围从 0 (参考黑) 到 +1 (参考白) 的抽象术语表示。

为了在参考观看条件下从参考显示器上得到预期的图像效果，需计算非线性三基色分量 R' 、 G' 、 B' ，见第 37 章。

历史上 NTSC 标准文件表明进行指数为 $1/2.2$ (大约为 0.45) 的“预校正”，但实际上使用 BT. 1886 EOCF。

在摄像机默认的上电状态下，非线性三基色分量是从摄像机的 RGB 三基色估计值按照前文描述的 BT. 709 OECF 光电转换函数计算得到的。该过程不严格地称为伽马校正。

显示基色的色彩度量性质应该符合 BT. 709 三基色标准：DTV 传输标准提倡用 BT. 709 标准，现代的消费级显示器使用 BT. 709 标准。然而，历史上 480i 内容的制作和加工使用的是 SMPTE 基本色度，而不是 BT. 709 标准（见表 26-4）。

39.6 亮度 (Y')

在 480i 系统中，亮度按非线性 $R'G'B'$ 基色分量的加权和计算，使用 BT. 601 的亮度系数，详见 28.6 节：

$${}^{601}Y' = 0.299R' + 0.587G' + 0.114B' \quad (39-1)$$

作为非线性 $R'G'B'$ 分量加权和的亮度分量 Y' 与色彩学中的 CIE 相对亮度 (Y) 没有简单的联系。视频编码规范通常不规定亮度带宽的上限（尽管传输标准可能有上限）。

39.7 图像中心、幅型比和消隐

图像中心位于 BT. 601 的 720 个有效采样点的中间两个点之间，在两个 0_H 时刻之间 321/572 的部分。有关垂直中心，如表 39-1 所示。

见 SMPTE RP 187《视频图像的中心、幅型比和消隐》。

在 4:3 系统中，幅型比定义为与净图像窗口像素点阵有关的 4:3，宽度为以 13.5MHz 速率采样的 708 个采样点，高度为 480 行。

31.2 节提到在一行的开始应避免从消隐到图像信息的瞬间转换。SMPTE 标准要求图像信息有 (140 ± 20) ns 的上升时间。对于 480i 或 576i 视频，消隐的转换最好用三段采样值实现，视频信号依次限制在其满幅度的 10%、50% 和 90%。

没有演播室标准处理 480i 视频的正方形采样的问题。作者建议使用 $780f_H$ 的采样率，即 $12\frac{3}{11}\text{MHz}$ (即 $12.272\overline{727}\text{MHz}$)。建议使用 648 个采样点，如果不行，也可用 644 或 640 个采样点，中心位置如上所述。

当有 480 或 512 个图像行的 MPEG-2 用于演播室时，底部图像对应第 525 行（见表 39-1）。底部左边平行（在第 263 行上）不在编码的图像行中。不幸的是，480i DV 系

统从此处开始还数字化了一行图像。

见 SMPTE RP 202 “MPEG-2 编码的视频对齐”。

39.8 半行消隐

大多数分量视频设备将两个场的顶部和底部行作为整体的行对待，半行消隐假定在转换到模拟信号时使用。在复合设备和模拟设备中，半行消隐一定要使用。

在复合域和模拟域中，在第 263 行左半行的图像底部视频信息应在 0_H 后的 $30.593\mu s$ 终止。这个时刻类似于一个完整行末期的消隐，不过是在两个 0_H 时刻的中点之前，而不是在 0_H 时刻本身之前。有

$$30.593 = \left(\frac{63.555}{2} - \frac{732 - 716}{13.5} \right)$$

历史上，在复合域和模拟域中，图像顶部的右半行，像是图像的第 284 行，起始于 0_H 后约 $41\mu s$ 。这个时刻类似于一个完整行开始的消隐，不过是在两个 0_H 时刻的中点之后，而不是在 0_H 时刻本身之后。然而，在 NTSC 广播中，第 284 行必须对隐匿字幕（与第 21 行一起）可用。因此，演播室设备再包含传统的第 284 行上的右半行图像没有意义。图像应被认为包含 482 个整行，加上第 263 行的左半行，有

$$41.259\mu s \approx \left(\frac{63.555}{2} + \frac{(858 - 732 + 2)}{13.5} \right) \mu s$$

半行消隐已经在逐行扫描视频以及 JPEG、MPEG 和高清中被废除。

39.9 数字分量 4 : 2 : 2 接口

数字视频的 C_B 和 C_R 色差分量通过缩放 $B'-Y'$ 和 $R'-Y'$ 分量得到，见 29.3 节的介绍。 $Y'C_B C_R$ 信号通过串行数字接口（SDI）传输。 $R'G'B'4 : 4 : 4$ （或 $R'G'B'A 4 : 4 : 4 : 4$ ）分量用可通过使用两个 SDI 通道的双链路接口传输；也可使用 SMPTE 344M 单链路 540Mb/s 的 SDI 接口传输。

在 480i 的 13.5MHz 采样中，位于 EAV 之后 16 个采样时钟间隔的采样点对应于行同步数据 (0_H)：如果被数字化，该采样值应该是模拟同步信号值的 50%。

图 39-2 显示了在 480i 分量数字 4 : 2 : 2 系统中的亮度波形。

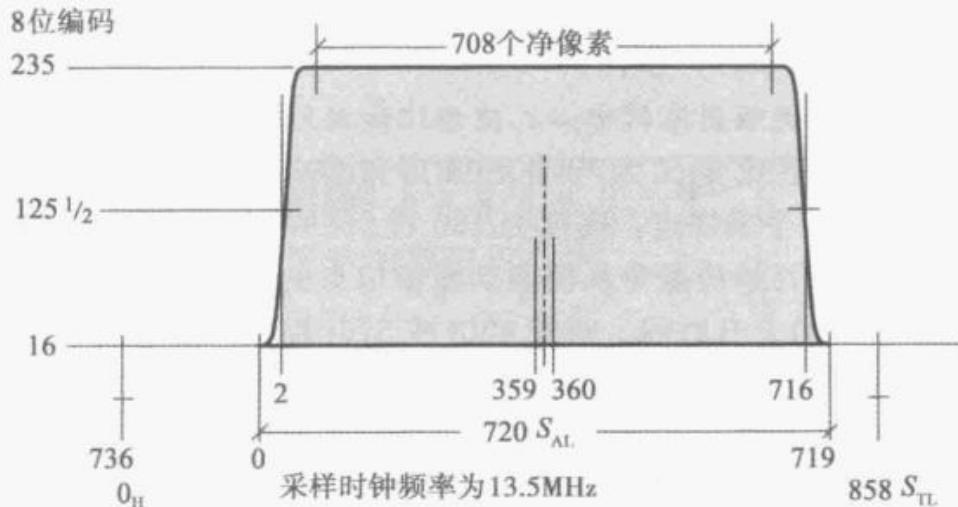


图 39-2 480i 数字分量 4 : 2 : 2 亮度波形

39.10 模拟分量 $R'G'B'$ 接口

分量模拟 480i $R'G'B'$ 接口是基于非线性 $R'G'B'$ 信号的。模拟 $R'G'B'$ 信号以电压形式传输，电压范围为从同步头到参考白色的 1V（允许瞬间的幅度稍高于参考白色）。

在演播室系统中，模拟分量 $R'G'B'$ 信号理想情况下建立电平为 0，所以式 (39-2) 的 0 对应于直流 0V。依照 SMPTE 253M (《三通道 RGB 模拟视频接口》)，其归一化值对应于 700mV。依据式 (39-2)，同步加到绿色分量上，sync 和 active 在被声明时为归一化值，否则为 0：

$$G'_{\text{sync}} = \frac{7}{10}(\text{active} \cdot G') + \frac{3}{10}(-\text{sync}) \quad (39-2)$$

遗憾的是，SMPTE $R'G'B'$ 模拟接口并不流行，“NTSC-相关的”电平更常用，有或没有建立电平。

有些系统，例如，日本的 480i 演播室视频，使用 10 : 4 的图像-同步比率以及 0 建立电平。在这种情况下，式 (39-3) 的归一化值对应 $(5/7)$ V，约为 714mV：

$$V'_{\text{sync}} = \frac{5}{7}(\text{active} \cdot V') + \frac{2}{7}(-\text{sync}) \quad (39-3)$$

很多系统，例如使用过时的 SIA RS-343-A 标准电平的计算机帧缓存，对视频分量进行编码类似于复合视频，使用 10 : 4 图像-同步比率，以及 7.5% 的建立电平：

$$V'_{\text{sync}} = \frac{3}{56}\text{active} + \frac{37}{56}(\text{active} \cdot V') + \frac{2}{7}(-\text{sync}) \quad (39-4)$$

39.11 EBU N10 模拟分量 $Y'P_BP_R$ 接口

模拟视频的 P_B 和 P_R 分量由 $B'-Y'$ 和 $R'-Y'$ 缩放而得到，在 29.2 节中有介绍。宽带 P_B 和 P_R 分量在理论上可行但很少使用；通常， P_B 和 P_R 经低通滤波减到亮度带宽的一半。

480i 中分量 $Y'P_BP_R$ 信号经常与电平符合 EBU Tech. N10 标准，建立电平为 0 的接口相连接。 Y' 、 P_B 和 P_R 的 0（参考消隐电平）对应于直流 0V 电平，归一化值对应于 700mV。同步加到亮度分量上，sync 在声明时为归一化值，否则为 0：

$$Y'_{\text{sync}} = \frac{7}{10}Y' + \frac{3}{10}(-\text{sync}) \quad (39-5)$$

图 39-3 显示了依据 EBC Tech. N10 标准的 480i 模拟分量接口的亮度波形。在北美，EBU N10 电平不可思议地公认为“SMPTE 电平”或“SMPTE/EBU N10 接口”，尽管

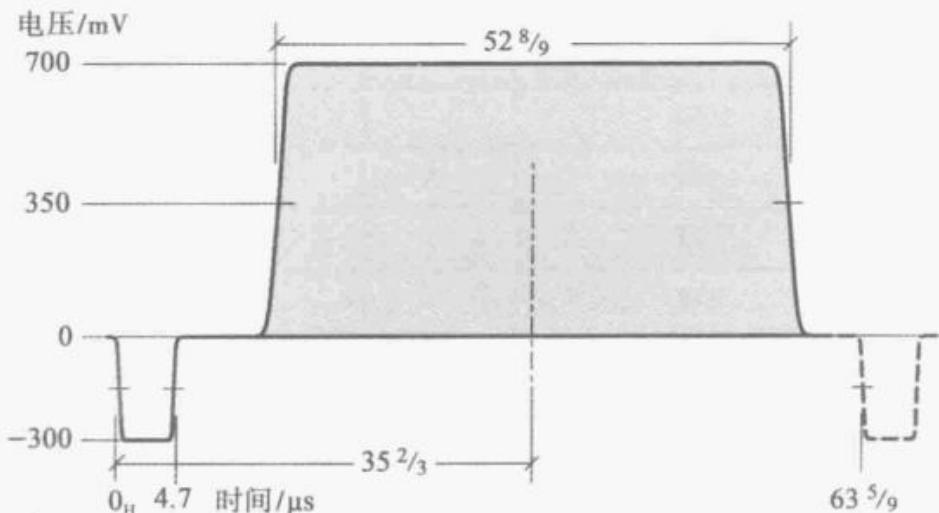


图 39-3 SMPTE 等级和零建立电平 480i 模拟分量亮度波形

N10 是一个独立的 EBU 标准，而 SMPTE 没能对分量模拟亮度/色差接口标准化。

CEA 对用于 DVD 播放器、机顶盒等消费电子设备的 700mV、零建立电平进行了标准化，用于幅型比为 4:3 和 16:9 的 480i 和 480p 格式。

见 EIA/CEA-770.2《标清电视模拟分量视频接口》。

2011 年，在引入模拟分量视频接口约四分之一世纪后，CEA（和定制电子设计与安装协会——CEDIA）标准化了接头使用的颜色：Y'用绿色，C_B用蓝色，C_R用红色（在消费电子领域，复合 NTSC 或 PAL 视频通常在有黄色接头的连线上传输）。

见 CEA/CEDAI-863-B“家庭影院系统连接的颜色编码”。

39.12 行业标准模拟分量 Y'P_BP_R 接口

不幸的是，在 SMPTE 就用于演播室的标准视频模拟分量接口达成一致之前，两家制造商的设备已经部署。尽管有时也作为可选项使用，SMPTE 标准极少用于 480i。相反，在使用的是两家“行业”标准：索尼和松下。理想的 Y'P_BP_R 命名法应当意味着亮度的建立电平为 0，色差分量和亮度有相同的偏移（从黑到白）。然而，这两个行业标准都使用建立电平，也没有一个让色差分量与亮度有相同的幅度。

索尼标清设备使用 10:4 的图像-同步比率（大约为 714mV 亮度和 286mV 同步），亮度的建立电平为 7.5%（图像的幅度范围为 $660 \frac{5}{7} \text{mV}$ ）。色差的范围是 $\pm 350\text{mV}$ 的 $4/3$ 倍，约为 $\pm 466 \frac{2}{3} \text{mV}$ (75% 的色条的 P_BP_R 范围是 $\pm 350\text{mV}$)。

松下标清设备使用 7:3 的图像-同步比率（大约为 700mV 亮度，300mV 同步），亮度的建立电平为 7.5%（图像的幅度范围为 647.5mV）。色差的缩放因子为 37/40，范围是 $\pm 323.75\text{mV}$ 。

第 40 章

576i 分量视频

本章详细介绍 576i25 (625/50/2 : 1) 视频的扫描、时序、同步结构和图像结构。本章介绍的扫描和时序信息适用于 576i25 视频所有的版本，包括模拟视频和数字视频。同步信息涉及分量模拟、模拟复合和复合数字系统。假定读者已熟悉 480i 分量视频。

40.1 帧频

576i 视频表示的是以时域上每秒 25 帧的恒定采样速率采样的静止或运动的二维图像。对于演播室视频，帧频的容差一般为 $\pm 4\text{ppm}$ 。实际中，误差适用于一个高频率的主时钟，但为了计算和标准的书写，引用帧频的公差比较方便。

40.2 隔行扫描

1 帧由 625 条相等时长的水平光栅线组成，按 2 : 1 隔行从上向下，从左向右地均匀扫描，形成第一场和第二场。第二场的扫描行相对于第一场的扫描行在垂直方向上有半个垂直采样间距的偏移，在时域上有半帧时间的滞后。用 MPEG-2 的术语表示，第一场就是顶场。

在一整帧中，各行从 1 开始连续编号。

表 40-1 显示了 576i 视频帧垂直结构，并标示出行号的分配和各行的内容。

表 40-1 576i 行分配

行号, 第一场 ($F=0$)	行号, 第一场 ($F=1$)	V	内容 (左半部分)	内容 (右半部分)
	313		EQ	BR
1		BR	BR	
	314		BR	BR
2			BR	BR
	315		BR	BR
3			BR	EQ
	316		EQ	EQ
4			EQ	EQ
	317		EQ	EQ
5			EQ	EQ
	318		EQ	无
6			垂直间隔视频	
	319		垂直间隔视频	
7~18			垂直间隔视频	
	320~331		垂直间隔视频	

(续)

行号, 第一场 ($F=0$)	行号, 第一场 ($F=1$)	V	内容 (左半部分)		内容 (右半部分)
19		$V=0$ (576 行)	VITC		
	332		VITC		
[†] 20			垂直间隔视频		
	[†] 333		垂直间隔视频		
[†] 21			VITC		
	[†] 334		VITC		
[†] 22			Quiet		
	[†] 335		Quiet		
23			WSS	图像	◊
	336~622		图像 (287 行)		
24~ [‡] 310			图像 (287 行)		
	[§] 623		图像	EQ	
311			EQ	EQ	
	624		EQ	EQ	
312			EQ	EQ	
	[¶] 625		EQ	EQ	

注: EQ 均衡脉冲。

BR 宽脉冲。

在 -135° 时抑制的突发信号。

§ 无条件抑制的突发信号。

¶ 在 $4f_{sc}$ PAL 中, 该行建议为 1137 个 S_{TL} (“复位”)。

† VANC 只允许在 20~22 行和 333~335 行出现。

◊ 右侧粗的垂直条表示根据 SMPTE RP 202 标准在 576i 或 576p MPEG-2 中承载的行 (图像的垂直中心位于 479 行和 167 行中间)。不幸的是, 576i DV 标准在此之前还量化了一个图像行。

从第 6 行到第 21 行, 以及从第 319 行到第 334 行可能包含与图像有关或无关的辅助 (垂直间隔) 信号。如果使用了垂直间隔时间码 (VITC), 冗余副本应置于第 19 (332) 行和第 21 (334) 行, 见 33.5 节。

关于 576i 中 VITC 的细节, 见 EBU 技术标准 N12, “电视录制的时序和控制代码”。

40.3 行同步

水平事件都是参照标记为 0_H 的时刻记录的。在模拟域, 0_H 定义为每个行同步脉冲的先导 (负跳变) 沿的 50% 点。在分量数字接口中, 同步与数字信息之间的对应关系由接口上传输的定时基准信号 (TRS) 来确定 (见 38.4 节)。

在模拟接口中, 每一行随着同步脉冲的负跳变沿开始于 0_H 。除了每场的垂直同步行之外, 每一行都起始于一个正常同步脉冲。起始于正常同步脉冲的每一行可能含有视频信息。除了被同步脉冲占用的间隔, 每一个起始于不是正常同步的同步脉冲的行保持消隐电平。

40.4 模拟场/帧同步

为了定义垂直同步, 帧分成半行周期的间隔。每个半行要么没有包含同步信息, 要么开始于同步脉冲出现。同步脉冲的持续时间为以下三种之一, 有 $\pm 0.100\mu s$ 的公差:

- 持续时间为 $4.7\mu s$ 的正常同步脉冲。

- 持续时间为正常同步脉冲一半的均衡脉冲。
- 持续时间为行时间的一半减去正常同步脉冲时间的宽脉冲，一帧中每一组 625 个半行都与以下垂直同步序列相关联：
 - 5 个前均衡脉冲。
 - 5 个宽脉冲。
 - 5 个后均衡脉冲。

第 623 行起始于一个正常同步脉冲并在行中间有一个均衡脉冲。第 318 行起始于均衡脉冲并保持消隐，中间没有同步脉冲。

模拟同步中，第 1 行和 0_V 定义为和 0_H 同时的第一个宽脉冲，见《Composite NTSC and PAL: Legacy Video Systems》第 2 章的表 2-3（不同于 480i 的习惯）。

图 40-1 显示了 576i 模拟视频的垂直同步结构。波形图是表 40-1 的模拟波形。该图显示了第一场和第二场的波形，给出了垂直同步间隔的细节。第一场有 312 行，第二场有 313 行。

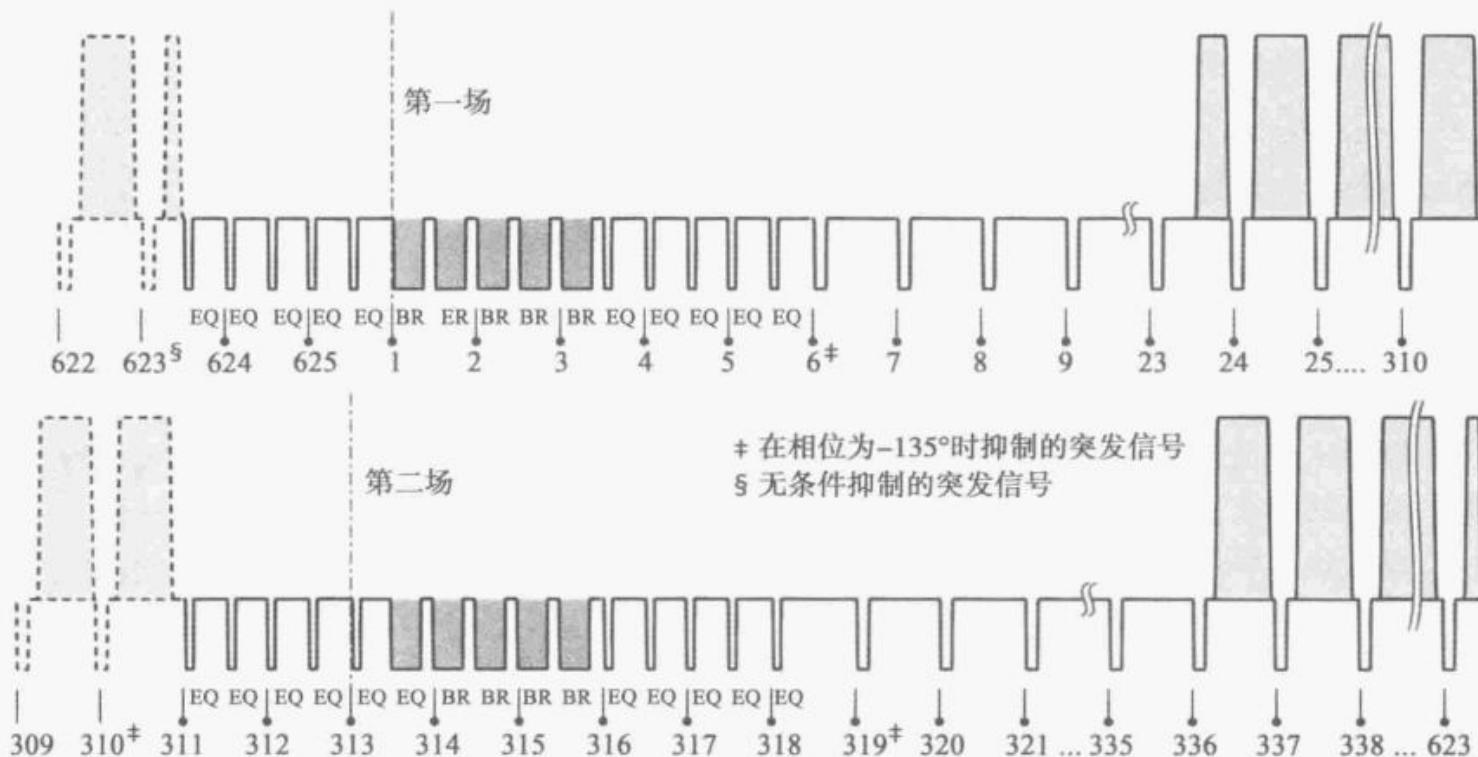


图 40-1 576i 垂直光栅

576i 系统同步与 480i 系统同步有几处差异。每场有 5 个预均衡脉冲、宽脉冲、过均衡脉冲，而不是 6 个。帧定义从包含图像最顶行的场开始，实际上是右半行（在 480i 扫描中，1 帧的第一个图像行是 1 整行，而图像顶部的右半行位于第二场）。

在 576i 系统中，行是从第 1 个宽同步脉冲开始编号的：预均衡脉冲是从一场结束而不是下一场开头进行计数的。可认为仅仅是命名的问题，但由于在数字视频接口中要对行序号编码，所以这是个实质性的问题。在 576i 系统中，行是在整个帧中编号的。

当同步表示为模拟或数字化的形式时，会有一个上升时间（10%~90%）为（200±20）ns 的升余弦转换沿，该转换沿的中点与理想的同步一致。

为了避免破坏同步或者图像，视频源间的切换在垂直间隔中进行。切换发生在每场第一个标准行的 0_H 的 $(30\pm 5)\mu s$ 后。在 576i 系统中，切换在第 6 行当中进行（如果第二场占主导，切换将在第 319 行当中进行）。

见 SMPTE RP 168 “同步视频切换的垂直间隔切换点的定义”。

40.5 $R'G'B'$ EOCF 和三基色

图像信息参照线性光基本的红、绿、蓝 (RGB) 三刺激值, 用范围从 0 (参考黑) 到 +1 (参考白) 的抽象术语表示。

为了在参考观看条件下从参考显示器上得到预期的图像效果 (见第 37 章), 需计算非线性三基色分量 R' 、 G' 、 B' 。

在摄像机默认的上电状态下, 非线性三基色分量是从摄像机的 RGB 三基色估计值按照 BT. 709 OECF 光电转换函数计算得到的。该过程不严格地称为伽马校正。

显示基色的色彩度量性质应该符合 BT. 709 三基色标准: DTV 传输标准提倡用 BT. 709 标准, 现代的消费级显示器使用 BT. 709 标准。然而, 历史上 576i 内容的制作和加工使用的是 SMPTE 基本色度, 而不是 BT. 709 标准 (见表 26-4)。

40.6 亮度 (Y')

在 576i 系统中, 亮度是根据 BT. 601 亮度系数计算非线性 $R'G'B'$ 基色分量的加权和得到的:

$${}^{601}Y' = 0.299R' + 0.587G' + 0.144B' \quad (40-1)$$

作为非线性 $R'G'B'$ 分量加权和的亮度分量 Y' 与色彩学中的 CIE 相对亮度 (Y) 没有简单的联系。视频编码规范通常不规定亮度带宽的上限 (尽管传输标准可能有上限)。

40.7 图像中心、幅型比和消隐

图像中心位于 BT. 601 的 720 个有效采样点的中间两个点之间, 在两个 0_H 时刻之间 983/1728 的部分。关于垂直中心, 如表 40-1 所示。

幅型比定义为 4 : 3 净图像窗口像素点阵, 690 个像素宽, 采样率为 13.5MHz, 566 行高。消隐的过渡不应该进入净图像窗口。

在复合及模拟域中, 第 623 行左半行的视频信息结束于 0_H 后 $(30.350 \pm 0.4) \mu\text{s}$ 。第 23 行右半行的视频信息开始于 0_H 后 $(42.500 \pm 0.1) \mu\text{s}$ 。

没有演播室标准处理 480i 视频的正方形采样的问题。作者建议使用 $944f_H$ 即 14.75MHz 的采样率。建议使用 768 个有效采样点, 中心位置如上所述。

当有 576 或 608 个图像行的 MPEG-2 用于演播室时, 底部图像行对应第 623 行 (见表 40-1)。底部左半行 (第 623 行) 也在编码的图像行中。当给到 MPEG 编码器时, 该行的右半行将为消隐; 在解码时, 可能有伪影。不幸的是, 576i DV 系统从此处开始还数字化了一行图像。

40.8 数字分量 4 : 2 : 2 接口

数字视频的 C_B 和 C_R 色差分量通过缩放 $B'-Y'$ 和 $R'-Y'$ 分量得到, 见 29.3 节的介绍。 $Y'C_B C_R$ 曾经通过 Rec. 656 和 EBU Tech. 3246 规定的并行数字接口传输, 现在用的是串行数字接口 (SDI)。

在 13.5MHz 采样率的 576i 系统中, 第 732 个采样点对应行同步基准 0_H 。若被数字化, 该采样值为模拟同步值的 50%。SMPTE RP 187 规定采样点 8 和 710 对应于图像宽度的 50% 处。对于平板显示, EBU 建议中央的 702 个采样点包含活动视频。

BT. 601 选择的 720 个有效采样点可同时满足 480i 和 576i 模拟视频消隐的需要: 720 个采样点足够容纳每个系统必需的过渡采样。

然而, 480i 和 576i 间消隐的公差不允许使用相同的消隐过渡采样点数: 480i 最窄的

可能图像宽度比 576i 的公差仍然宽出了几个采样点。

图 40-2 显示了 576i 数字分量 4:2:2 系统的亮度波形。

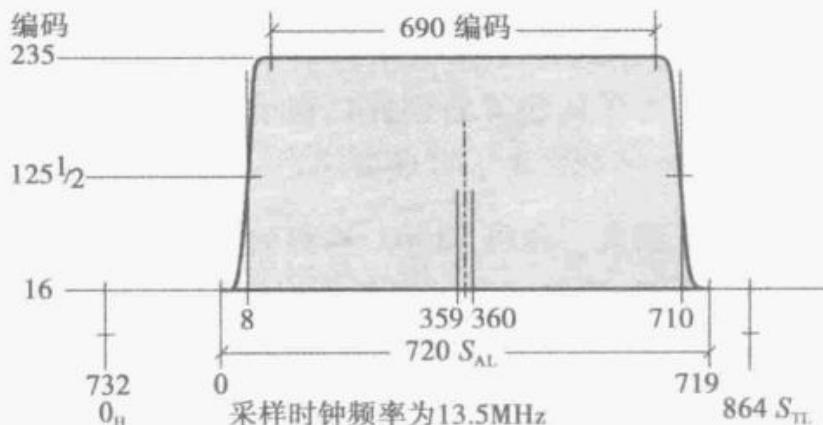


图 40-2 576i 数字分量 4:2:2 亮度波形

40.9 模拟分量 576i 接口

模拟分量 576i $R'G'B'$ 接口是基于以电压形式传输的非线性 $R'G'B'$ 信号的，电压范围为从同步头到参考白色的 1 V。允许瞬间的幅度略超出黑电平到白电平的范围。视频信号 0，即消隐电平，等于参考黑色，对应 0 V_{DC} 电平。视频信号的归一化值对应 700mV。

见 EBU Tech. N20《类似分量视频信号以 GRB 形式的并行接口》。

同步信号按下式加到绿色分量上，sync 和 active 在被声明时为归一化值，否则为 0：

$$G'_{\text{sync}} = \frac{7}{10}(\text{active} \cdot G') + \frac{3}{10}(-\text{sync}) \quad (40-2)$$

G' 信号从同步头到参考白色的幅度峰-峰值是 1V。576i 系统的电平单位常规定为 mV，而不是 480i 系统使用的 IRE 单位。若使用 IRE 单位，1 IRE 等于 7mV。

模拟亮度 (Y') 采用范围从同步头到参考白色峰-峰值为 1V 的电压传输。亮度信号 0，即消隐电平，等于参考黑色，对应直流 0V 电平。亮度信号的归一化值对应 700mV。同步信号按下式加到亮度分量上：

$$Y'_{\text{sync}} = \frac{7}{10}Y' + \frac{3}{10}(-\text{sync}) \quad (40-3)$$

见 EBU Tech. N1《模拟分量视频信号的并行接口》。

图像的 Y' 信号的行程为 700mV。图 40-3 显示了 576i 模拟分量接口中亮度的波形。

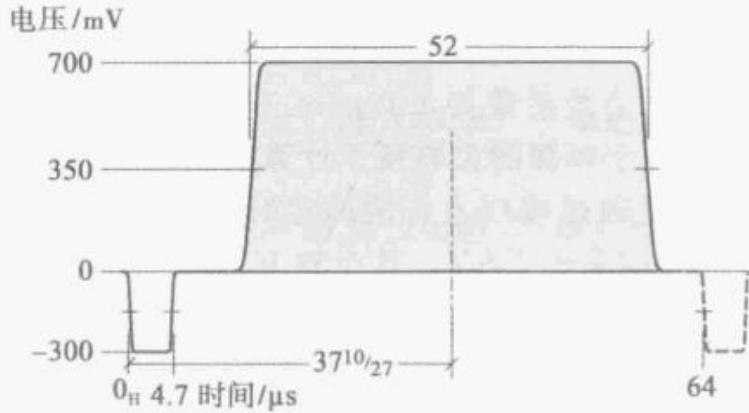


图 40-3 576i 模拟亮度分量的波形

P_B 和 P_R 色差分量由 $B'-Y'$ 和 $R'-Y'$ 缩放而得到，在 29.3 节中有介绍。尽管理论上可能有宽带 P_B 和 P_R 分量，实际中色差分量要经过低通滤波减到约亮度带宽的一半。 P_B 和 P_R 分量以电压形式传输，参考幅度为 ±350mV。

第 41 章

1280×720 像素高清

本章详细介绍 1280×720 像素（也称 720p）视频的扫描、时序和同步结构。本章的扫描和时序信息适用于所有的 720p 视频变体，包括模拟和数字。

见 SMPTE ST 296《1280×720 逐行图像采样结构——模拟和数字表示及模拟接口》。

41.1 扫描

720p 视频显示的是在时间上以 24/1.001、24、25、30/1.001、30、50、60/1.001 或 60 帧/s 的恒定速率采样的静止或运动的二维图像。采样率为 74.25MHz（在 720p59.94、720p29.97 和 720p23.976 中以 1000/1001 的因子修正）。所有这些系统均为 750 的总行数 (L_T)。为达到期望的帧频，每行的采样点数量 (S_{TL}) 有所调整。表 41-1 总结了扫描参数。

1 帧由总共 750 条等时长的水平光栅组成，均匀地从上到下，从左到右逐行扫描，从 1 开始连续编号。在总共 750 行中，有 720 行含有图像。表 41-2 给出了行号分配和各行的内容。

图像的垂直中心位于第 385 和第 386 行中间。

表 41-1 720p 扫描参数

系 统	f_s/MHz	S_{TL}
720p60	74.25	1650
720p59.94	74.25/1.001	1650
720p50	74.25	1980
720p30	74.25	3300
720p29.97	74.25/1.001	3300
720p25	74.25	3960
720p24	74.25	4125
720p23.976	74.25/1.001	4125

表 41-2 1280×720 行分配

行 号	内 容
1~5	三电平脉冲/宽脉冲（5 行）
6	消隐
7~25（19 行）	消隐/辅助
26~745（720 行）	图像〔净窗口 702 行〕
746~750（5 行）	消隐

在演播室视频中，帧频的公差正常为 $\pm 10\text{ppm}$ （百万分之一的相对偏差）。实际上公差应用于高频主时钟，但参照于帧频的公差便于计算和标准书写。

在数字接口中，视频信息通过接口上传输的定时基准信号 (TRS) 来识别（见 38.4 节）。一帧中最后一个有效行结束于 EAV，其中的 V 位为 1。该 EAV 标记了第 746 行的开始；下一帧的第一行开始于接下来的第 5 个 EAV。

41.2 模拟同步

水平事件参照于由三电平同步的过零点确定的 0_H 。数字采样值和模拟时序的关系是有效视频的第一个（编号为 0 的）采样值是在 0_H 时刻之后 260 个参考时钟间隔。

0_H 比 SAV 的第 1 个字提前 256 个时钟。 0_H 比 EAV 的第 1 个字落后 $S_{TL}-1280-260$ 个时钟。

在模拟接口中，每一行起始于一个三电平同步脉冲。三电平同步包括一个在 0_H 之前 40 个参考时钟间隔内为 (-300 ± 6) mV 的负电平，以及一个在 0_H 之后 40 个参考时钟间隔内为 $(+300 \pm 6)$ mV 的正电平。每个过渡沿的上升时间为 4 ± 1.5 个参考时钟间隔。

模拟域中垂直同步信号用宽脉冲表示，第 1 行到第 5 行各有一个。每个宽脉冲为 (-300 ± 6) mV，与活动视频即制作窗口的图像宽度时序相同。每个过渡沿的上升时间为 4 ± 1.5 个参考时钟间隔。第 1 行可当作帧的第 1 个宽脉冲来检测，也就是说，先是是没有宽脉冲的一行，后面是有宽脉冲的一行。

第 7 行到第 25 行没有承载图像信息，但可能承载了与图像有关或无关的辅助信号或其他信号。

模拟信号时序用数字标准来定义；数字采样频率定义了用于定义模拟时序的参考时间间隔。

图 41-1 显示了同步结构的细节，该波形图是表 41-2 的模拟形式。

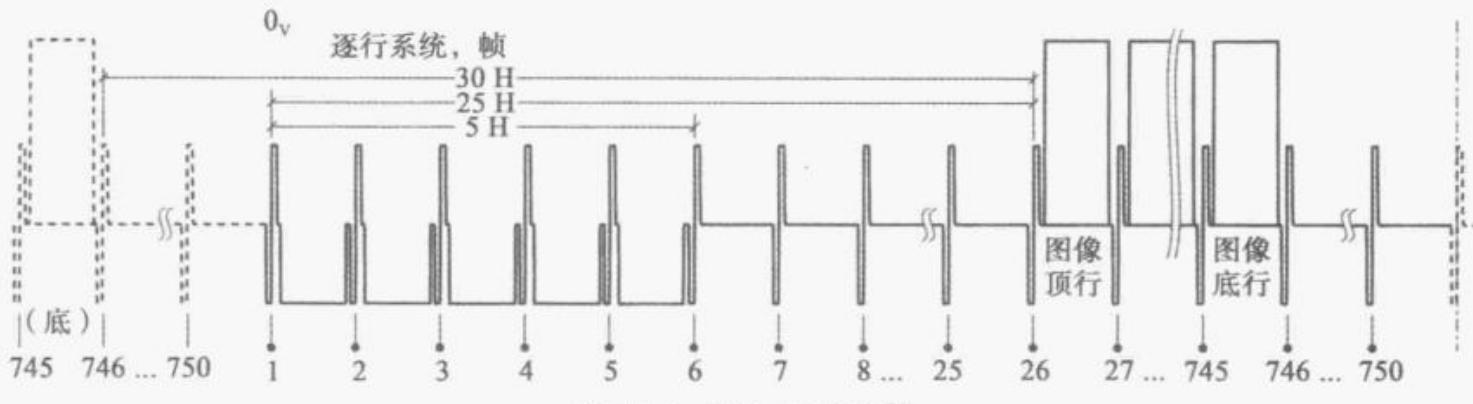


图 41-1 720p 垂直光栅

41.3 图像中心、幅型比和消隐

图像中心位于 1280 个有效采样点中央两个采样点的正中（即在第 639 和第 640 个采样点之间），并在 720 个图像行中央两行的正中（即第 385 行和第 386 行之间）。

幅型比相对于 1280×720 像素的制作窗口定义为 $16 : 9$ 。

31.2 节提到需要避免在一行开始时从消隐到图像信息的瞬间转换。宽为 1248 个采样点，高为 702 行的位于制作窗口中心的净图像窗口像素点阵应该保持在主观质量上不受边缘瞬变的影响。

41.4 $R'G'B'$ EOCF 和三基色

图像信息参照线性光基本的红、绿、蓝（RGB）三刺激值，用范围从 0（参考黑）到 +1（参考白）的抽象术语表示。

为了在参考观看条件下从参考显示器上（见第 37 章）得到预期的图像效果，需计算非线性三基色分量 R' 、 G' 、 B' 。

在摄像机默认的上电状态下，非线性三基色分量是从摄像机的 RGB 三基色估计值按照 BT. 709 OECF 光电转换函数计算得到。该过程不严格地称为伽马校正。

显示基色的色彩度量性质应该符合 BT. 709 三基色标准：DTV 传输标准提倡用 BT. 709 标准，现代的消费级显示器使用 BT. 709 标准。然而，在作者于 2011 年写作本书时，几乎所有的 50Hz 节目素材均是用 EBU 基色（见表 26-4）生成并加工的，几乎所有的 60Hz 节目素材均是用 SMPTE 基色（见表 26-5）生成并加工的。期望随着像是 OLED

等新的演播室显示技术的引入，形势能有所改观，但内容制作人和广播人的旧习惯很难改变。

41.5 亮度 (Y')

亮度是按 BT. 709 亮度系数计算的非线性 $R'G'B'$ 分量的加权和得到的：

$${}^{709}Y' = 0.2126R' + 0.7152G' + 0.0722B' \quad (41-1)$$

作为非线性 $R'G'B'$ 分量加权和的亮度分量 Y' 与色彩学中的 CIE 相对亮度 (Y) 没有简单的联系。高清的亮度公式不同于标清，见 28.10 节。视频编码规范通常不规定亮度带宽的上限。

41.6 数字分量 4：2：2 接口

$Y'C_B C_R$ 分量由 Y' 、 $B'-Y'$ 和 $R'-Y'$ 分量缩放而成，见 30.3 节。TRS 的插入见 38.4 节。HD-SDI 接口见 38.9 节。标准按 4：2：2 方案进行亚采样（见图 12-1 第三列）。如果采用 4：2：0 亚采样，即色差在垂直和水平方向上都进行亚采样，图像的质量也不会受到影响，但会给硬件设计和接口带来不便。

第42章

1920×1080 像素高清

本章详细介绍 SMPTE ST 274 标准化的一族高清晰度电视系统。这些系统的图像格式是 1920×1080 像素，幅型比为 16：9，正方形采样。

见 SMPTE ST 124《多种图像率的 1920×1080 扫描与模拟和并行数字接口》。

SMPTE ST 274 代表根据国际标准 BT.709 的色彩度量协议。先前的 SMPTE ST 274 版本在“临时”基础上允许使用 240M 颜色和亮度参数。制造商和用户需要用一些时间来完成向新标准的过渡。

基于 SMPTE 240M 的 1125/60 扫描和 1920×1035 像素图像格式、非正方形采样的高清设备已经部署了数年。该系统已经过时（该系统现在可看作 1080/30 或 1080/29.94 的 1035i 变体，有着 1035 个图像行而不是 1080 行，非正方形采样，色彩度量略有不同）。

42.1 扫描

1920×1080 像素视频显示的是在时间上以 24/1.001、24、25、30/1.001、30、50、60/1.001 或 60 帧/s 的恒定速率采样的静止或运动的二维图像。基础采样速率为 74.25MHz；对于非整数帧频的系统，采样率以 1000/1001 的因子修正。对于帧频高于 30Hz 的逐行扫描系统，采样率加倍到 148.5MHz（可能乘以 1000/1.001）。所有这些系统都有 1125 总行数 (L_T)，每行采样数 (S_{TL}) 根据所需的帧频来调整。表 42-1 总结了扫描参数，其中标记[¶]的 1080p 系统不允许在 ATSC 广播中使用。

表 42-1 1920×1080 像素格式扫描参数总结

系 统	f_S/MHz	S_{TL}
1080/p60 [¶]	148.5	2200
1080/p59.94 [¶]	148.5/1.001	2200
1080/p50 [¶]	148.5	2640
1080/30	74.25	2200
1080i29.97	74.25/1.001	2200
1080i25	74.25	2640
1080p30	74.25	2200
1080p29.97	74.25/1.001	2200
1080p25	74.25	2640
1080p24	74.25	2750
1080p23.976	74.25/1.001	2750

对于演播室视频，帧频的公差正常为±10ppm。实际上公差应用于高频主时钟，但为了计算和标准书写目的，公差参照于帧频比较方便。

1 帧包含共 1125 个等时长的水平光栅行，从上到下从左到右均匀扫描，从 1 开始顺序

编号。全部 1125 行中的 1080 行包含图像。表 42-2 说明了行号分配及各行的内容。

表 42-2 1080i 和 1080p 的行分配

行号 (连续的)	行号, 第一场 ($F=0$)	行号, 第二场 ($F=1$)	V	内容 (左半部分)	内容 (右半部分)
1	1	563			
2	2	564			
3	3	565			
4	4	566			
5	5	567			
6	6	568			
7~41 (35 行)	7~20 (14 行)			垂直间隔视频	
	569~583 (15 行)			垂直间隔视频	
42 ~ 1121 (1080 行)	21~560 (540 行)		$V=0$ (1080 行)	图像	
	584~1123 (540 行)			图像	
1122~1125 (4 行)	561~562 (2 行)	1124~1125 (2 行)			

逐行扫描系统按从上到下顺序每 1 帧传输 1080 个有效图像行。

隔行扫描系统先扫描 1 帧的第一场，再扫描第二场。每场的扫描行有每帧一半的垂直空间采样密度。第二场中的扫描行与第一场中的扫描行相比，在垂直方向差一个垂直采样间隔，在时间上晚半帧时间。第一场有 540 个有效图像行，从帧的顶部图像行开始。第二场有 540 个有效图像行，到帧的底部图像行结束。

在数字接口中，视频信息通过接口上传输的定时基准信号 (TRS) 来识别（见 38.4 节）。在逐行扫描系统中，1 帧的最后一个有效行结束于 V 位为 1 的 EAV。该 EAV 标记了第 1122 行的开始，下一帧的第 1 行开始于接下来的第 4 个 EAV。在隔行扫描系统中，一场的最后一个有效行结束于 V 位为 1 的 EAV。在第一场中，该 EAV 标记了第 561 行的开始。在第二场中，该 EAV 标记了第 1124 行的开始，下一帧的第 1 行开始于接下来的第二个 EAV。

42.2 模拟同步

在模拟接口中，每一行起始于一个三电平同步脉冲。三电平同步的过零点定义了行同步基准 0_H ，水平事件参照于 0_H 。数字采样值和模拟时序的关系是有效视频的第一个（编号为 0 的）采样值是在 0_H 时刻之后 192 个参考时钟间隔。

0_H 在 SAV 第 1 个字之前的 192 个时钟。 0_H 在 EAV 第 1 个字之后的 STL-1920-192 时钟。

三电平同步包括一个在 0_H 之前 44 个参考时钟间隔内为 $(-300 \pm 6) \text{ mV}$ 的负电平，以及一个在 0_H 之后 44 个参考时钟间隔内为 $(+300 \pm 6) \text{ mV}$ 的正电平。每个过渡沿的上升时间为 4 ± 1.5 个参考时钟间隔。

水平时序的细节如图 42-1 所示。图中的时间间隔以 74.25MHz 参考时钟的间隔表示，

0_H 定义为参考时钟的零时刻。为了得到有效视频中相对于 0_H 采样点的采样点编号，给这些计数值加上 $S_{AL} + 88$ （模为 S_{TL} ）。

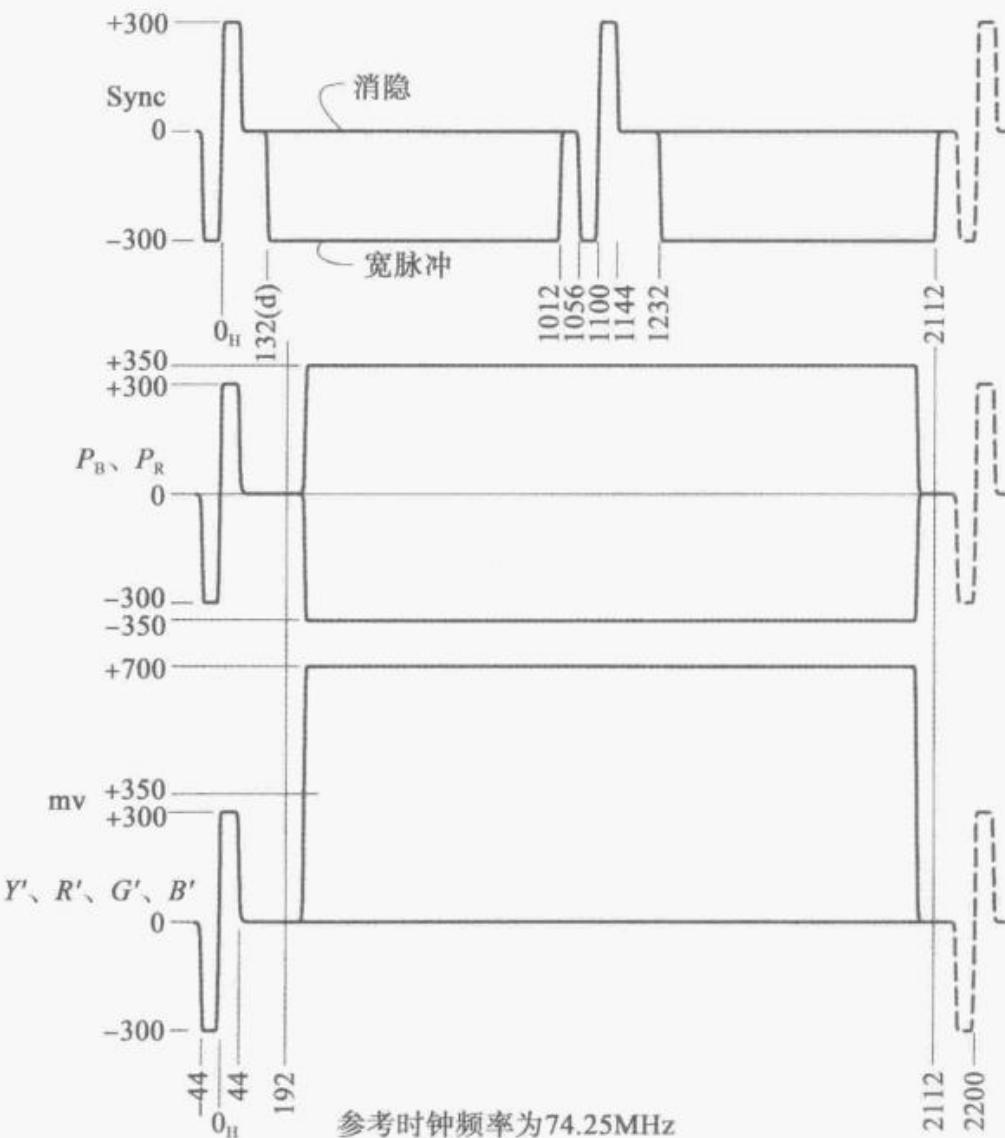


图 42-1 1080i30 模拟行的细节

模拟域中垂直同步信号用宽脉冲表示，逐行系统的宽脉冲结构与隔行系统的不一样。

逐行扫描系统每 1 帧有 5 个宽脉冲，从第 1 行到第 5 行各一个。每 1 个宽脉冲在 0_H 的 132 个参考时钟间隔之后声明为 (-300 ± 6) mV，在 0_H 的 2112 个参考时钟间隔之后撤销（宽脉冲的撤销与活动视频的结束，即制作窗口的右边沿时间相同）。每个过渡的上升时间为 4 ± 1.5 个参考时钟间隔。第 1 行由一帧的第一个宽脉冲定义，即先是没有任何宽脉冲的一行，后面是有宽脉冲的一行。第 6 行有第二个过零点位于 0_H 后的 $S_{TL}/2$ 个参考时钟间隔的三电平脉冲，该脉冲让人联想到模拟标清中的均衡脉冲。

在隔行扫描系统中，在垂直间隔的几行中有第二个过零点位于 0_H 后的 $S_{TL}/2$ 个参考时钟间隔的三电平脉冲。隔行扫描系统每场有 10 个宽脉冲，其排布如表 42-2 所示。每一个宽脉冲在前面紧挨着的三电平同步的过零点的 132 个参考时钟间隔，之后声明为 (-300 ± 6) mV，再过 880 个参考时钟间隔后撤销。每个过渡沿的上升时间为 4 ± 1.5 个参考时钟。第 1 行可作为左半行的第一个宽脉冲被解码，即检测一个正常行（没有宽脉冲和行中间的三电平同步）后面跟着一个 0_H 之后的宽脉冲。每一个宽脉冲的前面是一个过零点在 0_H 处或从 0_H 延迟 $S_{TL}/2$ 个参考时钟间隔处的三电平脉冲。

垂直间隔视频行不承载图像信息，可能承载与图像有关或无关的辅助信号和其他信号。

模拟信号时序用数字标准来定义；数字采样频率定义了用于定义模拟时序的参考时间间隔。

图 42-2 显示了垂直同步结构的细节；该波形图是表 42-2 所示的模拟形式。

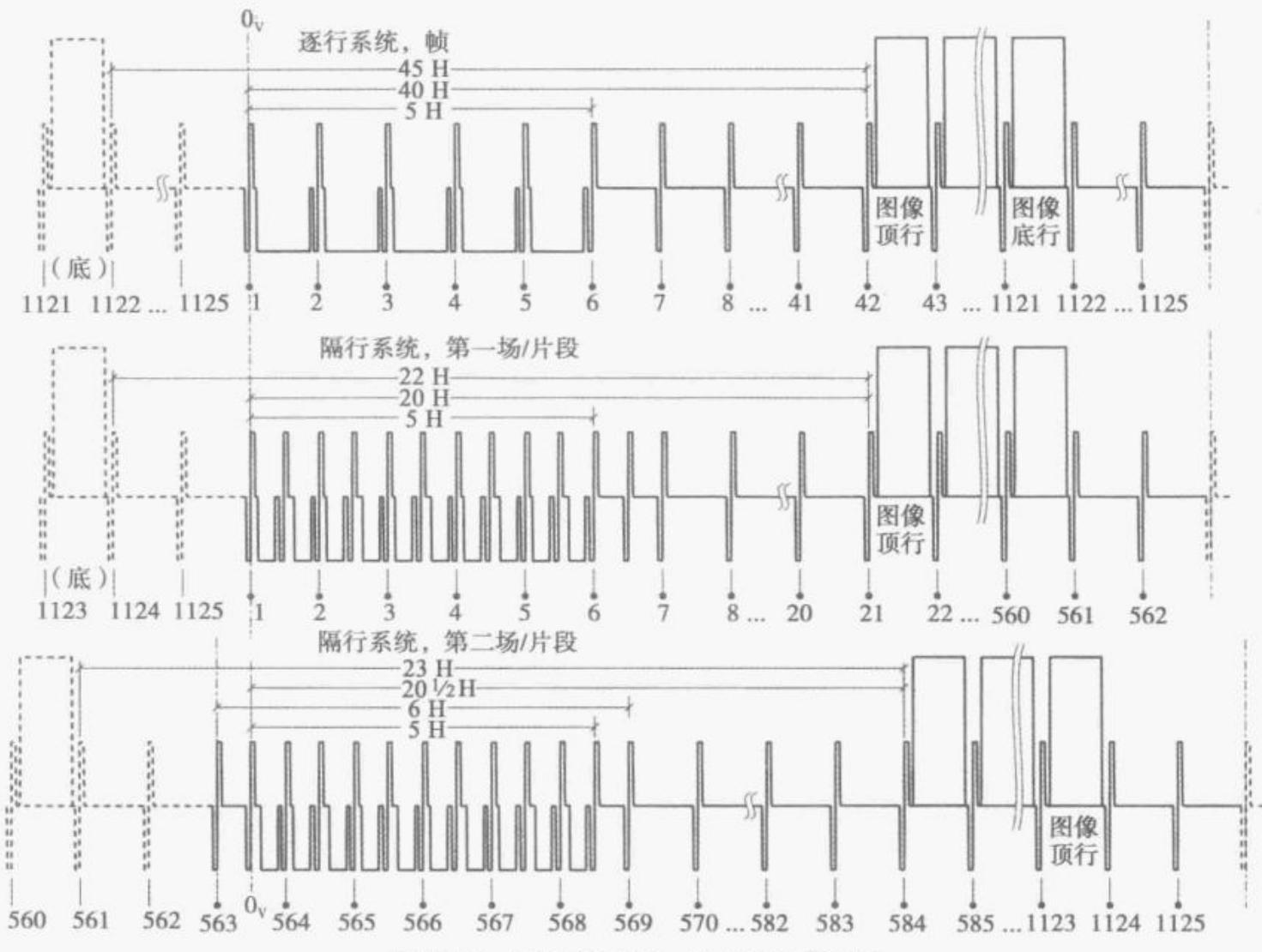


图 42-2 1080i 和 1080p 的垂直消隐间隔

42.3 图像中心、幅型比和消隐

图像中心位于 1920 个有效采样点中央两个采样点的正中（即采样点 959 和 960 之间），并在 1080 个图像行中央两行的正中（在逐行系统中是 581 行和 582 行之间，在隔行系统中是 290 行和 853 行之间）。

幅型比相对于 1920×1080 像素的制作窗口定义为 16 : 9。

在 31.2 节中提到需要避免在一行开始时从消隐到图像信息的瞬间转换。宽为 1888 个采样点，高为 1062 行的位于制作窗口中心的净图像窗口像素点阵应该保持在主观质量上不受边沿瞬变的影响。

42.4 $R'G'B'$ EOCF 和三基色

图像信息参照线性光基本的红、绿、蓝 (RGB) 三刺激值，用范围从 0 (参考黑) 到 +1 (参考白) 的抽象术语表示。

为了在参考观看条件下从参考显示器上（见第 37 章）得到预期的图像效果，需计算非线性三基色分量 R' 、 G' 、 B' 。

在摄像机默认的上电状态下，非线性三基色分量是从摄像机的 RGB 三基色估计值按

照 BT. 709 OECF 的光电转换函数计算得到。该过程不严格地称为伽马校正。

显示基色的色彩度量性质应该符合 BT. 709 三基色标准：DTV 传输标准提倡用 BT. 709 标准，现代的消费级显示器使用 BT. 709 标准。然而，在作者于 2011 年写作本书时，几乎所有的 50Hz 节目素材均是用 EBU 基色生成并加工的，几乎所有的 60Hz 节目素材均是用 SMPTE 基色生成并加工的。期望随着像是 OLED 等新的演播室显示技术的引入形势能有所改观，但内容制作人和广播人的旧习惯很难改变。

42.5 亮度 (Y')

亮度是按“BT. 709 亮度”介绍的亮度系数计算的非线性 $R'G'B'$ 分量的加权和得到的：

$${}^{709}Y' = 0.2126R' + 0.7152G' + 0.0722B' \quad (42-1)$$

作为非线性 $R'G'B'$ 分量加权和的亮度分量 Y' 与色彩学中的 CIE 相对亮度 (Y) 没有简单的联系。高清的亮度公式不同于标清，见 28.10 节。视频编码规范通常不规定亮度带宽的上限。

42.6 数字分量 4:2:2 接口

$Y'C_B C_R$ 分量由 Y' 、 $B'-Y'$ 和 $R'-Y'$ 分量缩放而成，见 30.3 节。HD-SDI 接口见 38.9 节。模拟接口详见 38.10 节。

第43章

高清录像带

由于本书第一版出版以来，录像带的使用大大减少。大量老式的标清录像带设备仍然在使用，但是现在硬盘驱动器和闪存介质很容易满足标清的需求。录像带仍然相当广泛地应用于高清视频，比起硬盘驱动器和闪存介质仍有一些优势。另外，几种最初为录像带而设计的压缩系统现在已经成功转变为使用硬盘驱动器和闪存介质。

表 43-1 总结了高清数字录像带的格式。D-6 格式是非压缩的（尽管有色度亚采样），所有其他格式均进行了压缩。

表 43-1 高清数字录像带格式

标记	方法	磁带	数据传输速率 / (Mb/s)	注释
D-5 HD (HD-D5, D-15)	分量压缩, 4 : 2 : 2 高清	12.65mm (1/2in, 源于 VHS)	270	(Panasonic)
D-6	未压缩的 4 : 2 : 2 高清	19mm	1188	
HDCAM (D-11)	分量压缩的, 类似 M-JPEG, 1440 × 1080, 3 : 1 : 1 高清	12.7mm (源于 Beta)	135	(Sony)
DVCPRO HD (D-12)	分量压缩的, 1280 × 1080 (or 960 × 720), 4 : 2 : 2 高清	6.35mm	100	DV100
HDCAM SR (D-16)	分量压缩的, MPEG-4 SSIP, 1920 × 1080, 4 : 2 : 2 (Lite) 4 : 2 : 2 (SQ) 4 : 4 : 4 (HQ)	12.65mm (源于 Beta)	220 440 880	(Sony)

43.1 D-5 HD (HD-D5、D-15)

SMPTE D-5 标准定义了以 270Mb/s 数据传输速率录制非压缩标清内容。松下给 D-5 格式配备了压缩比约为 5 : 1 的 M-JPEG 编解码器，以适应高清内容，该编码与 DV 编码很相似。这种 D-5 VTR 的变体标记为 D-5 HD 格式，或 HD-D5 格式，之后 SMPTE 将其标准化为 D-15 格式。

D-5 HD 系统以 59.94Hz 或 60.00Hz 录制 720p 或 1080i 视频，进行 4 : 2 : 2 色度亚采样。

D-5 HD 压缩与 DV 的一个差异是使用了重叠变换，8×8 块中最右边一列采样值与右方的下一个块的最左边一列重叠。解码时，冗余编码的列由两个相邻块的适当的采样值求平均来重建。与 DV 中使用的非重叠变换相比，这种方案减小了块效应。

见 Malvar, Henrique S. (1992 年)《Signal Processing with Lapped Transforms》(Norwood, Mass. : Artech House)。

43.2 D-6

SMPTE D-6 定义了数据传输速率为 1.188Gb/s 的非压缩 8 位 4：2：2 高清录像带格式。这种设备性能卓越且非常昂贵，现在已淘汰。

43.3 HDCAM (D-11)

HDCAM 是由索尼开发和商业化的。具有几种帧频中任何一种帧频的隔行或逐行高清视频降采样到 1440×1080 像素图像格式，并进行 3：1：1 色度亚采样。采用 M-JPEG 风格的压缩方式，与 DV 的压缩方式相当，使用重叠变换（像是 D-5 HD），编码速率约为 135Mb/s。HDCAM 使用的盒式录像带源自 Betacam 的 1/2 英寸磁带。这种录像带录制方案标准化为 SMPTE D-11。

1920×1080 像素到 1440×1080 像素的降采样意味着亮度采样的幅型比相对于原始亮度阵列为 4/3（色度亚采样等效为 4：1：1）。

43.4 DVCPRO HD (D-12)

DV25/DV50 标准被调整为适合 100Mb/s 以容纳降采样到 1280×1080 像素图像格式的 1080i30 高清信号（或 720p60 降采样到 960×720 像素），在降采样域色度进行 4：2：2 亚采样。这种压缩方案记作 DV100，有时也称为 DVCPRO100；数字磁带录像机由松下作为 DVCPRO HD 推出，分享了许多 DV25 和 DV50 数字录像机的机械和信号处理元素。这种录像带录制方案之后被标准化为 SMPTE D-12。

1920×1080 像素到 1280×1080 像素的降采样意味着亮度采样的幅型比相对于原始亮度阵列为 3/2（色度亚采样等效为 3：1：1）。

43.5 HDCAM SR (D-16)

HDCAM SR 的压缩符合 MPEG-4 第 2 部分“简单演播室类 (SStP)”。HDCAM SR 以 440Mb/s 速率 $Y'C_B C_R$ 4：2：2 (SQ 模式) 或 880Mb/s 速率 $R'G'B'$ 4：4：4 (HQ 模式) 录制 1920×1080 像素视频。HDCAM SR 设备也可在 880Mb/s 模式下录制或播放两路独立的 $Y'C_B C_R$ 4：2：2 码流。增加的一种附加模式 SR 精简版的数据传输速率为 220Mb/s (“精简版”指数据速率需求，而压缩工作是繁重的)。

第44章

模拟分量高清接口

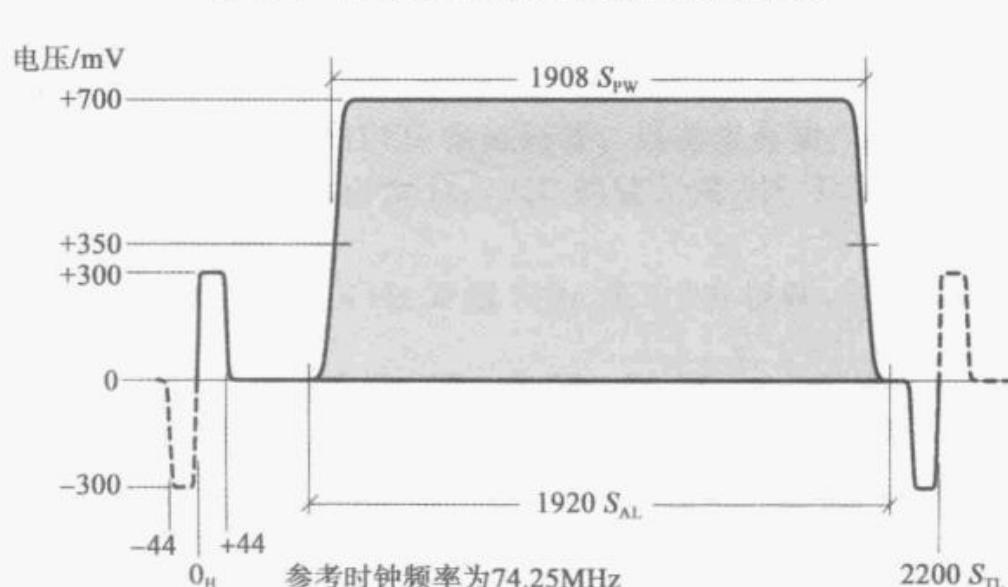
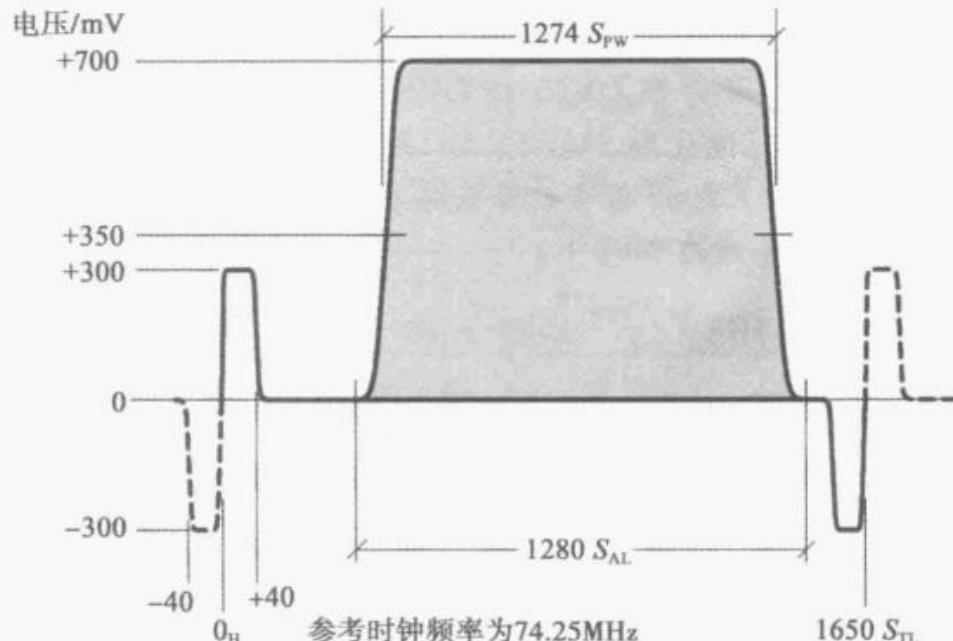
在模拟域中，模拟分量高清信号在接口上作为 $Y'P_BP_R$ 信号以电压的形式传输。亮度的范围从同步头到参考白色峰-峰值为 1V。亮度信号 0 即消隐电平，等于参考黑色，对应直流 0V。归一化的亮度信号对应 700mV。允许信号幅度在短时间内稍微超出参考黑色到参考白色的范围。

同步信号根据式 (44-1) 加到亮度信号上，sync 和 active 在声明时取归一化值，否则取 0：

$$Y'_{\text{sync}} = \frac{7}{10}Y' + \frac{3}{10}(-\text{sync}) \quad (44-1)$$

Y' 信号从同步头到参考白色的幅度峰-峰值为 1V。一个 IRE 单位对应 7mV。

图 44-1 显示了 720p60 模拟分量接口的亮度波形；图 44-2 显示了 1080i30 中模拟分量接口的亮度波形。图中的时间间隔为 74.25MHz 参考时钟的间隔。



P_B 和 P_R 色差分量由缩放 $B'-Y'$ 和 $R'-Y'$ 分量而得，见 30.2 节的介绍。尽管理论上可能有宽带 P_B 和 P_R 分量，但实际上色差分量要经低通滤波得到约一半亮度带宽。 P_B 和 P_R 分量以电压形式传输，参考幅度为 $\pm 350\text{mV}$ 。

SMPTE 标准没有规定模拟接口中 R' 、 G' 和 B' 分量的时间重合度的公差。作者建议各分量在时间上相互重合，与同步信号的误差在 $1/4$ 个采样时钟内，对 720p60 和 1080i30 来说，时间重合度是 $\pm 3.4\text{ns}$ 。

EIA/CEA 已经将这里描述的 720p60 和 1080i30 系统标准化，包括这些 $Y'P_BP_R$ 电平，用于消费类高清电子设备。

见 EIA/CEA-770.3。

前滤波和后滤波特性

720p59.94、720p60 或 1080i30 的 Y' 、 R' 、 G' 、 B' 分量信号有 30MHz 的标称通带，色差分量有 15MHz 的标称通带。

图 44-3 描绘了对 R' 、 G' 、 B' 和 Y' 分量信号进行预滤波和后滤波的滤波器特性。模拟 P_B 和 P_R 色差分量信号使用在频率轴上除以 2 的相同模板进行前滤波和后滤波。这些特性是对 BT. 601 中的模板进行频率缩放而得的，有少量修改。

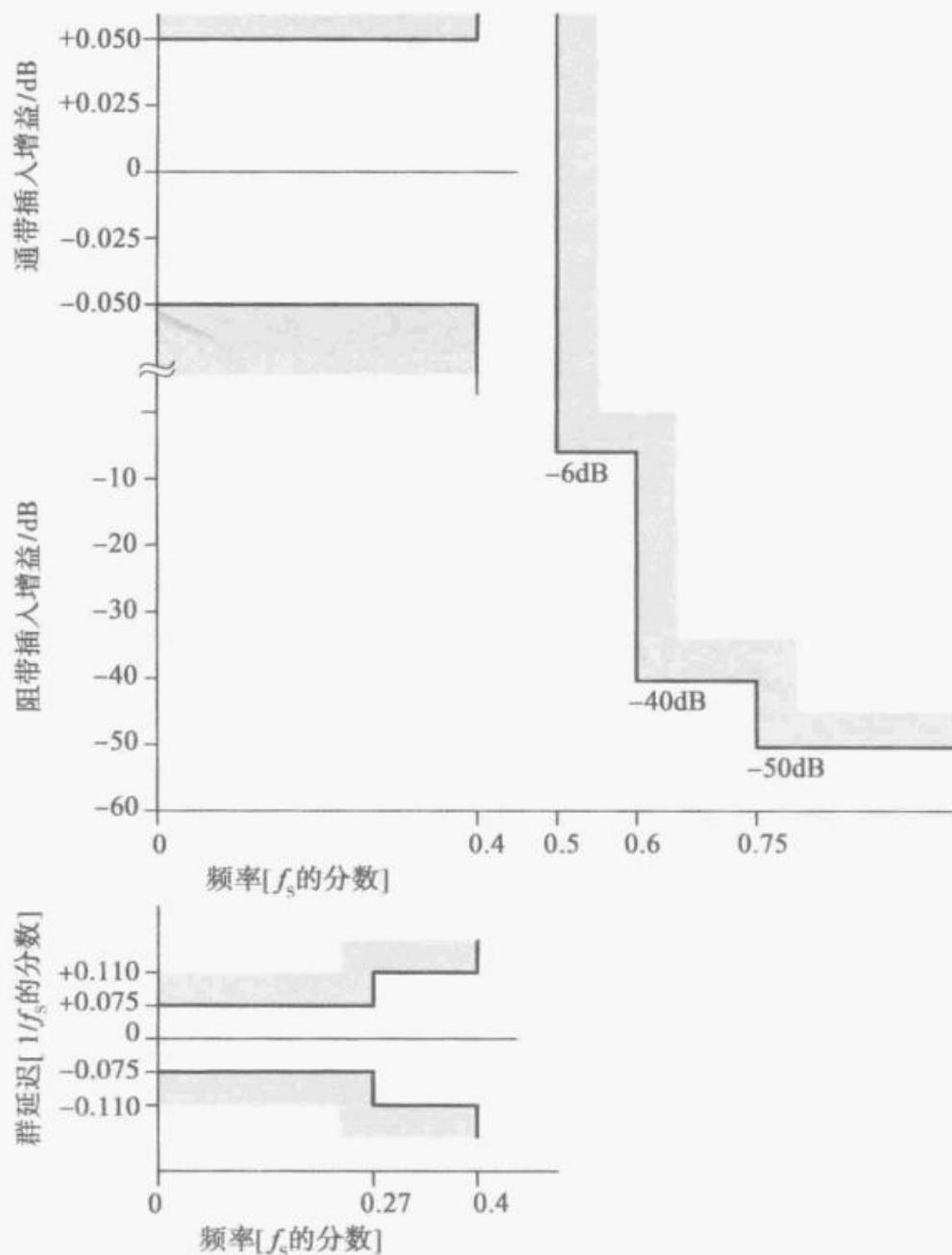


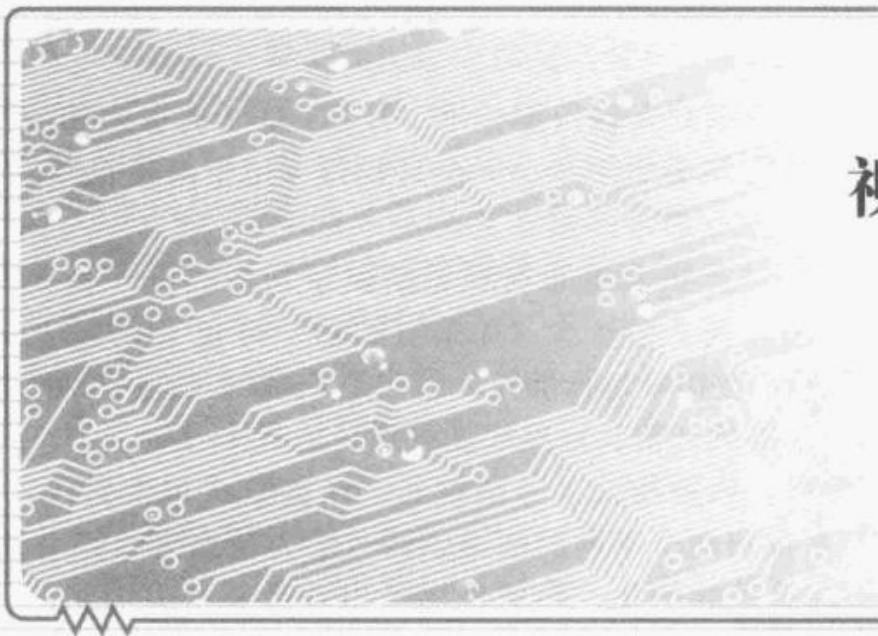
图 44-3 Y' 和 $R'G'B'$ 分量的滤波器模板

通带幅度纹波的公差相对于在 100kHz 的插入损耗为 $\pm 0.05\text{dB}$ 。在 Y' 、 R' 、 G' 和 B' 分量一半采样率时的插入损耗为 6dB 或更大。

SMPTE ST 274 在其“附录 B 信息性”中提供了亮度和色差模板。在标准术语里，“信息性”这个词说明该信息随标准一起发布，但声明符合标准并不需要遵从附录内容。

第五部分

视频压缩



第 45 章

JPEG 和 M-JPEG 压缩

本章介绍 JPEG，一种静态图像的有损压缩标准。JPEG 基于离散余弦变换（DCT）。JPEG 很少直接用于视频，但它是 M-JPEG（motion-JPEG）（用于桌面视频编辑）和 DV 压缩的基础。JPEG 技术也是 MPEG 技术的核心。

M-JPEG 指使用类似 JPEG 的算法来压缩视频场或帧序列中的每一场或每一帧的技术。M-JPEG 系统使用 JPEG 的方法，但极少（如果有的话）符合 ISO/IEC JPEG 标准。DV 是一种特定类型的 M-JPEG，被很好地标准化，将在第 46 章中描述。MPEG-2 的仅用 I 帧的变体在概念上和 M-JPEG 等效，但其标准很受尊重，见第 47 章。

JPEG 标准定义了 4 种模式：顺序模式、分层模式、渐进模式和无损模式。JPEG 标准可容纳 2 到 16 位的 DCT 系数和两个不同的熵编码器（哈夫曼编码和算术编码）。基准指的是 JPEG 顺序模式的一个规定的子集，限制使用 8 位系数和哈夫曼编码。只有基准 JPEG 在商业上具重要性，JPEG 的其他模式主要被学术界关注，这里不做讨论。

见 ISO/IEC 10918 “信息技术——连续色调静止图像的数字压缩与编码”。

45.1 JPEG 块和 MCU

一个 8×8 采样数据阵列在 JPEG 术语中称为块。在进行彩色图像的 JPEG 压缩之前，通常非线性 $R'G'B'$ 数据经矩阵变换为 $Y'C_B C_R$ ，接着进行 $4:2:0$ 亚采样。根据 JPEG 标准（以及将要介绍的 JFIF 标准），其他彩色亚采样方案也是允许的；奇怪的是，允许 C_B 、 C_R 的亚采样比率不一样。然而，只有 $4:2:0$ 被广泛应用，所以剩余这部分的假定采用了 $4:2:0$ 。4 个 8×8 亮度块、1 个 $8 \times 8 C_B$ 块和 1 个 $8 \times 8 C_R$ 块在 JPEG 术语中称为最小编码单元（MCU）。MCU 对应 DV 或 MPEG 术语中的宏块。图 45-1 显示了 $4:2:0$ 宏块排列。

MCU 包含 6 个 8×8 块：4 个亮度块，一个 C_B 块和一个 C_R 块。6 个组成块是非线性 $R'G'B'$ 经矩阵运算得到 $Y'C_B C_R$ ，后再按 $4:2:0$ 方案亚采样的结果；色度亚采样实际是压缩的第一阶段。各个块相互独立处理。

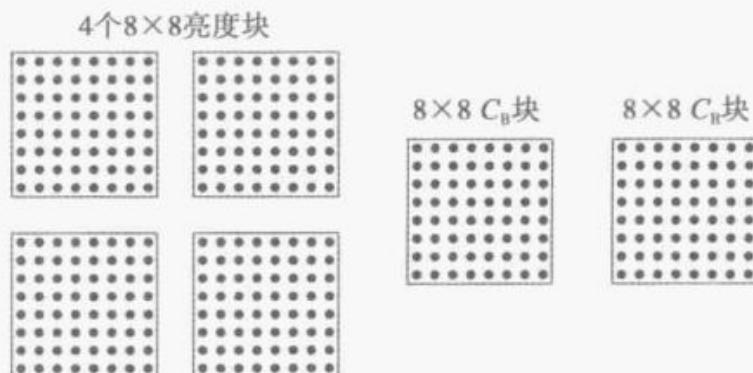


图 45-1 JPEG 4:2:0 最小编码单元（MCU）

在 MPEG 中，一个宏块是被一个 16×16 亮度采样点阵列覆盖的区域。在 DV 中，一个宏块包括被 8×8 色度采样点阵列（块）覆盖的 Y' 、 C_B 和 C_R 块。在 JPEG 中，一个 MCU 包含被最小片的 Y' 、 C_B 和 C_R 块覆盖的那些块。对于 $4:2:0$ 亚采样，所有这些定

义都是等效的，但对于 4:1:1 和 4:2:2（或其他很少使用的 JPEG 样式）来说不同。

对于桌面图形，保存高质量的 JPEG 可能使 $R'G'B'$ 通道（分量）各自被压缩而不经过亚采样。

JPEG 对亮度块和色差块用几乎相同的算法各自独立处理。唯一的显著区别是色度块和亮度块使用的量化矩阵和可变长编码（VLE）表通常不一样。

正如 22.1 节的解释，典型的图像由低空间频率的能量主导。图 45-4 显示了图像中一个 8×8 亮度采样值矩阵的例子。图 45-2 显示了该亮度矩阵经 DCT 得到的 8×8 空间频率矩阵。矩阵左上角的 [0, 0] 项（直流项）表示零频率能量，该项通常包含很大的值，就没在这里画出。[0, 0] 附近的系数倾向于有较高的值，距离 [0, 0] 更远的系数倾向于值变小。取决于图像数据，几个孤立的高频系数可能有很高的值。

这种空间频率域中图像能量的典型分布表示图像中存在冗余。通过在空间频率域中编码而不是直接对图像采样值编码可减少冗余。

除了从典型图像数据中消除冗余的优势，用空间频率表示还有另一项优势。视觉系统的光线敏感度取决于空间频率：从图 23-5 可以看出，人对低空间频率比对高空间频率更为敏感。高空间频率上的信息可大大退化而不会给图像质量带来令人不快的（也可能甚至是能感觉到的）影响。一旦图像数据通过 DCT 变换，高阶系数可以取近似值，即粗量化，以丢弃那些对应于对图像的可感知质量贡献很小的空间频率分量的数据。

原则上，DCT 算法可用于任意的块尺寸，从 2×2 到整幅图像大小，也许是 512×512 （DCT 在应用于一个维数是 2 的幂次的矩阵时效率最高）。DCT 的视频应用中选择 8×8 亮度块代表了在足够小的块可以最小化存储和处理开销，而足够大的块可以有效地利用图像冗余之间的折中。

DCT 运算丢弃视觉不敏感的图像信息。不过，出乎意料的是，JPEG 标准本身并未涉及感知一致性。因为 JPEG 的目标是重现视觉上重要的信息，所以呈现给 JPEG 算法的所谓 RGB 值先要经过模仿视觉的非线性变换，例如 1.5 节概述的方法。

45.2 JPEG 框图

如图 45-3 中的 JPEG 框图所示，在图的顶部是 JPEG 编码器的三个主要模块：离散余弦变换（DCT）运算（有时称为前向 DCT，即 FDCT）、量化（Q）和可变长编码（VLE）。解码器（图 45-3 底部）则以相反的顺序执行这些操作的每一步逆操作。反离散余弦变换有时记作 IDCT；反量化有时称为去量化，有时记作 IQ。

由于 8 行高的垂直变换，编码器的 DCT 子系统和解码器的 IDCT (DCT^{-1})

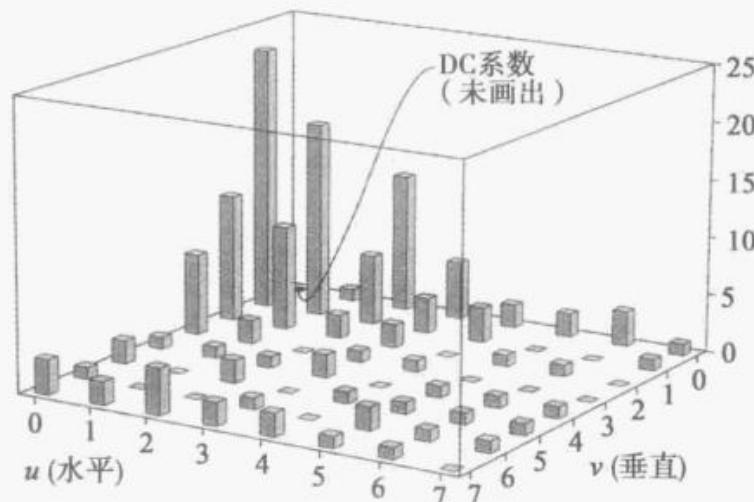


图 45-2 DCT 变换集中图像能量

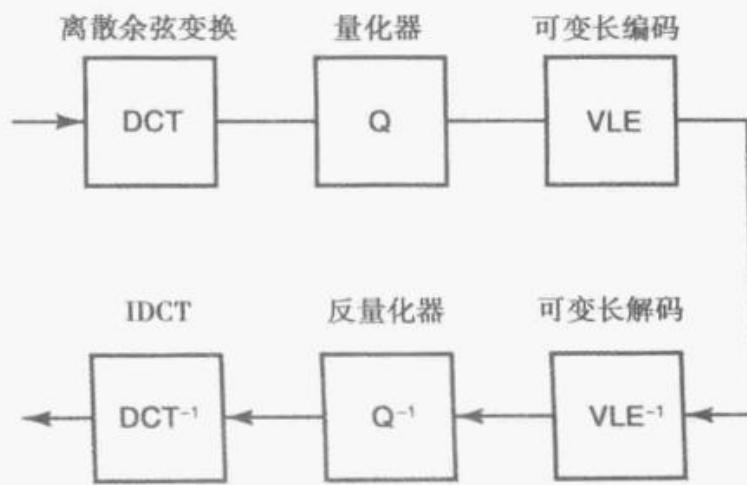


图 45-3 JPEG 框图

子系统中需要 8 行的图像存储器。当在多数情况下 DCT 以可分离的形式实现时，8 行的图像存储器称为转置存储器。

45.3 电平转换

JPEG 的 DCT 公式适用于有符号的采样值。在普通的硬件或固件中，DCT 是以定点、二进制补码运算实现的。标准视频接口使用偏移二进制表示，所以每一个亮度或色差采样值要在 DCT 前减去 2^{k-1} 进行电平转换，其中 k 为使用的位数。

45.4 离散余弦变换

8×8 前向离散余弦变换 (FDCT) 输入一个有 64 个采样值的 8×8 数组 (记作 f , 其元素为 $f_{i,j}$)，产生一个有 64 个变换系数的 8×8 数组 (记作 F , 其元素为 $F_{u,v}$)。FDCT 可表示为：

$$F_{u,v} = \frac{1}{4} C(u) C(v) \sum_{i=1}^7 \sum_{j=0}^7 f_{i,j} \cos\left[\frac{(2i+1)u\pi}{16}\right] \cos\left[\frac{(2j+1)v\pi}{16}\right]$$

$$C(w) = \begin{cases} \frac{1}{\sqrt{2}} & (w = 0) \\ 1 & (w = 1, 2, \dots, 7) \end{cases} \quad (45-1)$$

余弦项不需要当时计算，可以预先算好存储在表格中。

反变换 IDCT (DCT⁻¹) 可表示为：

$$f_{i,j} = \frac{1}{4} \sum_{u=0}^7 \sum_{v=0}^7 C(u) C(v) F_{u,v} \cos\left[\frac{(2i+1)u\pi}{116}\right] \cos\left[\frac{(2j+1)v\pi}{16}\right] \quad (45-2)$$

前向和反向变换涉及几乎相同的算法：编码和解码的复杂度非常相似。IDCT 是自身的反变换（用一个缩放因子），所以对变换系数执行 DCT 可完美重建原始样值，只受 DCT 和 IDCT 舍入误差的影响。

如果直接按以上等式实现，一个 8×8 DCT 为求出 64 个结果系数中的每一个系数需要 64 次乘法运算（和 49 次加法运算），总共 4096 次乘法运算，平均每个像素需 8 次乘法运算。然而，DCT 是可分离的：一个 8×8 DCT 可以按 8 个 8×1 水平变换后接着 8 个 1×8 垂直变换来计算。这种优化，结合其他类似快速傅里叶变换 (FFT) 的优化，很大程度上减小了计算复杂度：一个完全优化的 8×8 DCT 对于每 8 个采样值（或 IDCT 中的变换系数）仅需 11 次乘法运算。

45.5 JPEG 编码示例

这里通过演算一个数值化的例子来介绍 JPEG 编码。图 45-4 表示图像中一个在电平转换之前的 8×8 亮度样值数组。

计算 DCT 并舍入为整型的结果如图 45-5 所示。

该例说明了图像能量集中至低频变换系数，即在 DCT 矩阵左上角的那些系数。这个阶段没有信息丢失。DCT 是自身的反变换，所以进行第二次 DCT 能完美重建原始采样值，只受 DCT 和 IDCT 的舍入误差影响。

如式 45-1 所示，一个 8×8 DCT 运算实际上使系数值相对于原始采样值乘以因子 8。 $[0, 0]$ 项（即直流系数或直流项）的值 1260 是原始采样值总和的 $1/8$ （所有其他系数均为交流系数）。

$$f = \begin{bmatrix} 139 & 144 & 149 & 163 & 155 & 155 & 155 & 155 \\ 144 & 151 & 153 & 156 & 159 & 156 & 156 & 156 \\ 150 & 155 & 160 & 163 & 158 & 156 & 156 & 156 \\ 159 & 161 & 162 & 160 & 160 & 159 & 159 & 159 \\ 159 & 160 & 161 & 162 & 162 & 155 & 155 & 155 \\ 161 & 161 & 161 & 161 & 160 & 157 & 157 & 157 \\ 162 & 162 & 161 & 163 & 162 & 157 & 157 & 157 \\ 162 & 162 & 161 & 161 & 163 & 158 & 158 & 158 \end{bmatrix}$$

图 45-4 图像的 8×8 亮度样值数组

$$F = \begin{bmatrix} 1260 & -1 & -12 & -5 & 2 & -2 & -3 & 1 \\ -23 & -17 & -6 & -3 & -3 & 0 & 0 & 1 \\ -11 & -9 & -2 & 2 & 0 & -1 & -1 & 0 \\ -7 & -2 & 0 & 1 & 1 & 0 & 0 & 0 \\ -1 & -1 & 1 & 2 & 0 & -1 & 1 & 1 \\ 2 & 0 & 2 & 0 & -1 & 1 & 1 & -1 \\ -1 & 0 & 0 & -1 & 0 & 2 & 1 & -1 \\ -3 & 2 & -4 & -2 & 2 & 1 & -1 & 0 \end{bmatrix}$$

图 45-5 DCT 将图像块的能量集中于低频 DCT 系数

人类视觉系统对高空间频率上的信息不是很敏感。在不引入可察觉的失真的情况下，高空间频率上的信息可以在一定程度上被丢弃。JPEG 使用一个量化矩阵 (Q)，为 64 个空间频率分别编码了量化步长。在 JPEG 压缩的量化步骤中，每个变换系数除以 Q 矩阵中对应的量化器值（量化步长）。除法后的余数（分数）被舍弃。

不是 DCT 本身，而是在变换系数量化后舍弃分数使得 JPEG 是有损的！

图 45-6 给出了一个 Q 矩阵示例。 Q 矩阵各项形成了图 23-5 所示的一个放射状对称版本。量化矩阵中 $[0, 0]$ 项相对较小（在这里是 16），所以直流 (DC) 项被精细量化。离 $[0, 0]$ 项越远，各项变得越大，量化变得越粗糙。由于对应高阶系数的步长大，所以可用更少的位数来表示高阶系数。

在 JPEG 和 MPEG 标准以及大多数类似 JPEG 的方案中，每个量化矩阵项取值从 1 到 255。

$$Q = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

图 45-6 典型的 JPEG 量化矩阵

乍看起来，与 DC 系数相关联的大步长（此处 $Q_{0,0} = 16$ ）让人不放心：从 -127 到 128 的 8 位数据，由于被 16 除，你可能预计量化后的系数可以只用 4 位表示。然而，如前所述，式 (45-1) 的算术运算相对于采样值将系数放大了 8 倍，所以相对于采样值，量化器值为 16 时对应 7 位精度。

图 45-7 给出了量化后的离散余弦变换 (DCT) 系数和舍去的商分数。

块中大多数高频信息，即矩阵右边和底部的 DCT 项量化为 0。除了直流项，非零系数的幅度很小。

在量化之后，量化后的系数根据块中图像能量的可能分布重新排列。这是通过图 45-8 所示的之字形 (zigzag) 扫描来完成的。

$$F' = \begin{bmatrix} 79 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

图 45-7 量化后的 DCT 系数

$$F' = \begin{bmatrix} 79 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ -2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

图 45-8 之字形扫描

一经重新排列，量化后的系数就用一个一维字符串表示；块结束符 (EOB) 标记着字

字符串中其后所有系数都为零，如图 45-9 所示。

79	0	-2	-1	-1	-1	0	0	-1	EOB
----	---	----	----	----	----	---	---	----	-----

图 45-9 之字形扫描后的系数字符串

通常情况下只有几个高阶的量化后的系数非零，之字形重排倾向于产生连 0 的字符串。额外的压缩用可变长编码（VLE，又称哈夫曼编码）来实现。可变长编码是一个无损过程，它利用了变换后系数的“游程”（零码的计数）和“电平”（绝对值或幅度）的统计特性。

在 JPEG 和 MPEG 术语中，系数幅度（绝对值）称为电平。

直流项的处理比较特殊，使用了差分编码。第一个直流项直接编码（利用直流 VLE 表），但后面的直流项作为与第一项的差编码。实质上，前一个直流项用作当前项的预测器。 Y' 、 C_B 和 C_R 维持独立的预测器。

零交流系数被压缩，字符串用〈游程，电平〉对来表示，如图 45-10 所示。

{1: -2}, {0: -1}, {0: -1}, {0: -1}, {2: -1}, EOB
--

图 45-10 VLE 〈游程，电平〉对

JPEG 编码器有一个或多个 VLE 表，将〈游程，电平〉对的集合映射为可变长位串；高概率的对被分配短的位串。JPEG 没有标准的 VLE 表。然而，经常使用非标准附录给出的范例 VLE 表。通常有两个表，一个用于亮度，一个用于色差。图像使用的表包含在 JPEG 位流的头部，从而传送给解码器。

45.6 JPEG 解码

以相反的顺序执行编码器操作的逆操作就实现了解压缩。图 45-11 显示了本例中原始采样值和重建采样值的差异矩阵，即重建误差。原始采样值不能完美地被重建是 JPEG 的典型情况。然而，根据人类视觉的空间频率响应来丢弃信息，确保了压缩过程引入的误差不太容易感觉到。

JPEG 的性能可以不严格地用原始图像数据和重建数据的误差表示。像是均方误差（MSE）这样的度量将测量变得客观，但是 MSE（以及其他工程和数学测量）不一定和主观性能很契合。实际上，根据感知属性来选择量化矩阵。JPEG 解压缩后不完美恢复的原始图像数据实际是在图像上叠加了噪声。直流项的不完美重建导致 JPEG 的 8×8 块变得可见，即出现 JPEG 块伪影。

JPEG 失真及其压缩几乎全部来自量化步骤。DCT 自身可能引入少量的舍入误差，IDCT 也可能引入轻微的舍入误差。可变长编码和解码过程是无损的。

45.7 压缩率控制

量化矩阵中的项越大，压缩率越高。JPEG 的压缩率控制可通过改变量化矩阵实现，图 45-12 所示的手动控制即表示这个意思。量化步长越大，压缩率越高，但如果步长太大，图像质量会受影响。步长越小，质量越好，但这以压缩比降低为代价。没有一个简单

$$\boldsymbol{\varepsilon} = \begin{bmatrix} -5 & -2 & 0 & 1 & 1 & -1 & -1 & -1 \\ -4 & 1 & 1 & 2 & 3 & 0 & 0 & 0 \\ -5 & -1 & 3 & 5 & 0 & -1 & 0 & 1 \\ -1 & 0 & 1 & -2 & -1 & 0 & 2 & 4 \\ -4 & -3 & -3 & -1 & 0 & -5 & -3 & -1 \\ -2 & -2 & -3 & -3 & -2 & -3 & -1 & 0 \\ 2 & 1 & -1 & 1 & 0 & -4 & -2 & -1 \\ 4 & 3 & 0 & 0 & 1 & -3 & -1 & 0 \end{bmatrix}$$

图 45-11 重建误差

的方法能在真正进行压缩之前预测对一幅特定的图像压缩后的数据有多少字节。

原则上可以对一幅特定图像自适应地选择能使性能最优的量化矩阵。然而，这并不实际。用于静态图像的 JPEG 编码器一般提供几种压缩设置的选项，每种设置与一个系统设计师选中的固定量化器相关联。

在 ISO JPEG 中，量化矩阵直接在位流中传输。在 JPEG 的 DV 改编版本中，标准定义了几个量化矩阵，位流表明要使用哪一个。

由于不同图像可能关联不同的量化矩阵，量化矩阵必须传给解码器，如图 45-13 所示，要么作为文件的一部分，要么经由一个辅助通道。在彩色图像中，通常亮度分量和色度分量的量化器是分开的。在静态帧应用中，这个操作的开销很小。在实时系统中，每一帧或每一帧之内都传输量化矩阵的开销是一种负担。

一种改进的压缩率控制方法被多种 M-JPEG 格式采用（也包括在 45.9 节将看到的 M-PPEG）：建立起参考亮度和色度量化矩阵，矩阵的全部项用一个数值化的参数进行缩放，该参数为量化器缩放因子（QSF，有时标记为 Mquant）。改变 QSF 可实现压缩率控制。

前面提到 JPEG 通常使用 4:2:0 色度亚采样的亮度/色度编码。不过，JPEG 标准允许未经亚采样的 $R'G'B'$ 图像数据，以及四通道的图像数据（例如，用于打印的 CMYK）。这些方案不适用于视频。

45.8 JPEG/JFIF

ISO/IEC JPEG 标准定义了位流，与 JPEG 将跨越各种通信链路使用的初衷一致。为了将 JPEG 技术应用于计算机文件，需要少量的增补信息；此外，需要将 ISO/IEC 位流编码成字节流。单图像 JPEG 文件的事实标准是 JPEG 文件交换格式（JFIF），被 C-Cube 领导的行业组织所采纳。一个 JFIF 文件封装了 JPEG 位流，以及少量的补充数据。若有一个描述为 JPEG 的图像数据文件，它几乎肯定是一个 JFIF 封装的 ISO/IEC JPEG 位流。

见 Hamilton, Eric (1992),《JPEG 文件交换格式》版本 1.02 (加利福尼亚州 Milpitas: C-Cube Microsystems)。这份非正式文件获 ECMA 认可，并于 2009 年 6 月作为 ECMA TR/98 出版，具有相同标题。

JPEG 标准本身暗示 JPEG 也可应用于线性光 RGB 数据。可是 JPEG 视觉特性不佳，除非应用于基于感知编码的图像数据，即伽马校正后的 $R'G'B'$ 。

ISO/IEC JPEG 标准本身似乎暗示该技术可以应用于任意的 RGB 数据。标准本身没有提到基本色度、白色点、传递函数和伽马校正。为了获得准确的颜色，还必须采用在标准之外的手段传输这些参数。在 Mac OS X 10.6 之前，有两类 JPEG/JFIF 文件，PC 和 Macintosh。在经典 Macintosh 上创建的文件符合默认的 Macintosh 编码，其 $R'G'B'$ 编码



图 45-12 JPEG 的压缩率控制

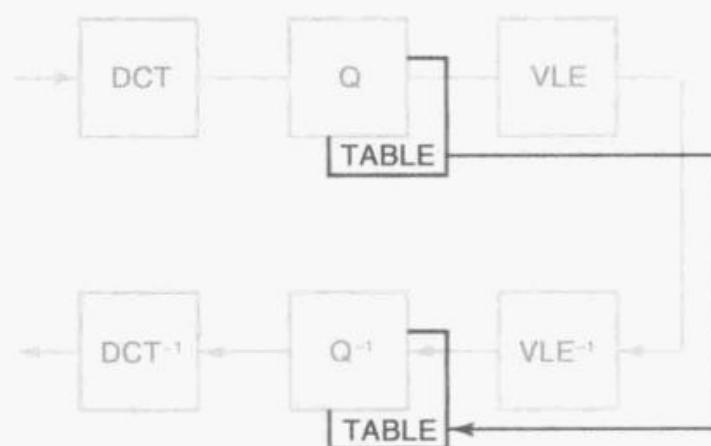


图 45-13 量化矩阵通过辅助通道传输给解码器

应当提升到 1.52 次幂函数来产生显示器三刺激色值。没有可靠的方案来区分这两类文件。在 PC 上和现代 Mac 上创建的文件均解释为 sRGB 编码。

45.9 M-JPEG

M-JPEG (Motion-JPEG) 指使用类似 JPEG 的算法来压缩视频场或帧序列中的每一幅图片的技术。“类似 JPEG”的意思是使用的算法具有由 JPEG 标准化算法的所有一般特征，包括 DCT、量化、之字形扫描和可变长编码的算法。然而，通常不会生成 ISO JPEG 位流，而且一些系统加入了 JPEG 标准以外的算法特性。各种 M-JPEG 系统在桌面视频编辑中广泛使用，但没有很好确立的标准，并且压缩的视频文件通常不能在 M-JPEG 系统之间互换。

在演播室应用中，文件互换是一种实际需要，现在已有两种方法。这两种方法都在功能上等价于 M-JPEG，但是都有稳固的标准。

第一种方法是 DV 压缩，为录像带上的消费级数字录制而开发，不过也用于桌面视频编辑。DV 压缩将在第 46 章介绍。

第二种方法是 MPEG-2 视频压缩，在第 47 章介绍。MPEG-2 利用帧间的一致性实现比 M-JPEG 高得多的压缩率，并计划主要用于视频分发。然而，MPEG-2 的只有 I 图像（有时也称为 I 帧）的版本在功能上等价于 M-JPEG，正用于演播室编辑。

有几种演播室 DVTR 格式，像数字 Betacam 和 HD-D5，使用 M-JPEG 风格的压缩，但不与任何 JPEG、DV 和 MPEG 有密切联系。

延伸阅读

Clarke 阐述了静态图像的传输编码理论。Rabbani 和 Jones 写了一篇很好的对静态图像压缩数学理论的介绍。Symes 提供了对视频压缩通俗易懂的介绍。

CLARKE, R. J. (1985), *Transform Coding of Images* (Boston: Academic Press) .

RABBANI, MAJID, and PAUL W. JONES (1991), *Digital Image Compression Techniques* (Bellingham, Wah.: SPIE) .

SYMES, PETER (2003), *Digital Video Compression* (New York: McGraw-Hill) .

第46章

DV 压缩

DV 是指推出时针对消费类数字视频磁带 (DVC) 录像机, 后来为专业应用而加以修改的压缩和数据打包方案。DV 压缩采用离散余弦变换 (DCT)、量化和可变长编码 (VLE), 与 JPEG 类似; 不过 DV 不符合 JPEG 标准: 为适应隔行扫描、恒定数据传输速率 (CBR) 操作, 以及其他与磁带录像相关的特性而做了优化。隔行是通过允许编码器在帧模式和场模式之间进行动态选择来处理的。恒定数据传输速率是通过动态改变量化矩阵以避免超出可用容量来实现的。

消费类 DVC 具有 25Mb/s 的数据传输速率, 作者称之为 DV25。DV25 编码被 D-7 (DVCPRO) 和 DVCAM 采纳用于演播室应用; 它扩展到 50Mb/s (DV50), 用于 DVCPRO50 和 D-9 系统; 然后扩展到 100Mb/s (DV100), 用于 D-11 (DVCPRO HD)。

DV25 和 DV50 压缩 BT. 601 480i29.97 或 576i25 标清视频。DV100 压缩 BT. 709 高清视频。DV 的色度亚采样五花八门, 如表 46-1 所示。

表 46-1 DV 色度亚采样

	480i	576i	HD
DV25 (消费类, DVCAM)	4 : 1 : 1	4 : 2 : 0	
DV25 (DVCPRO, D-7)	4 : 1 : 1	4 : 1 : 1	
DV50 (DVCPRO50, D-9)	4 : 2 : 2	4 : 2 : 2	
DV100 (DVCPRO HD, D-12)			4 : 2 : 2

46.1 DV 色度亚采样

不同版本的 DV 使用不同的亚采样方案: 4 : 2 : 2、4 : 1 : 1、4 : 2 : 0, 甚至 3 : 1 : 1 和 3 : 1 : 0! 在 DV 中, 宏块由一个 C_B 采样值的 8×8 块、一个 C_R 采样值的 8×8 块, 以及规定数量与排列的亮度采样值的 8×8 块组成。

SMPTE 314M 声明在从 4 : 2 : 2 到 4 : 1 : 1 亚采样时“每隔一个像素丢弃一个”。显然, 高图像品质需要在丢弃采样值前进行适当的滤波。在 DV 中, C_B 和 C_R 与亮度在水平与垂直方向上均重合。然而, 在 576i 消费类设备使用的 4 : 2 : 0 方案中, C_R 样点与 C_B 样点不在同一位置。相反, 在每一场中 C_B 和 C_R 样点处在逐行交替的垂直位置上: 每个 C_B 样点位于相关的 C_R 样点的两个图像行之下。

DV 格式起源于消费类应用。消费类 DV25 把 27 个宏块聚集成一个超级块 (SB), 覆盖 48 个图像行, 排列如图 46-1 所示。在 480i29.97 DV25 中, 每帧有 50 个超级块; 在 576i25 DV25 中, 每帧有 60 个超级块。

图 46-1 显示了三种色度亚采样方案的超级块——4 : 2 : 2 (顶上), 4 : 2 : 0 (中间), 4 : 1 : 1 (底下)。细线围起来的是宏块。每个宏块包含一个 $8 \times 8 C_B$ 块、一个 $8 \times 8 C_R$ 块和两个或更多的 8×8 亮度块。粗线围起来的是超级块; 每个超级块由 27 个宏块组成。一个 4 : 1 : 1 宏块的宽是高的 4 倍。

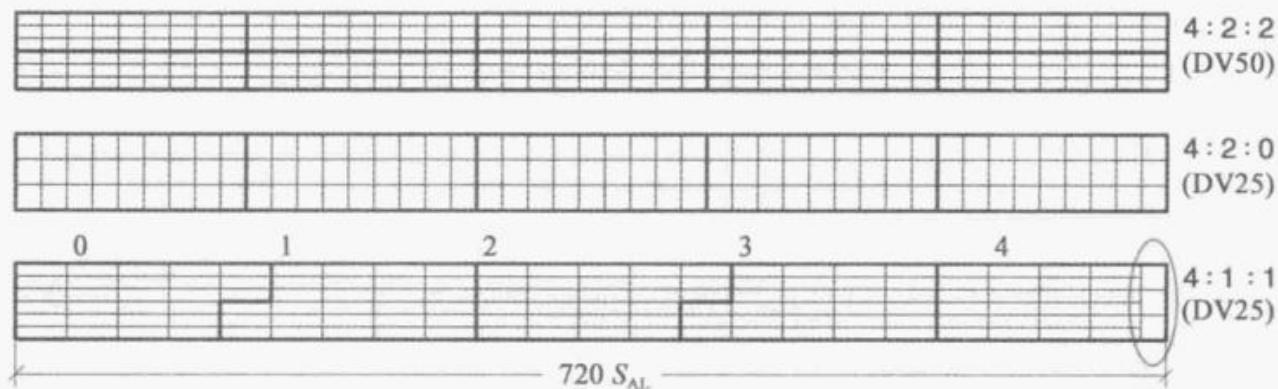


图 46-1 DV 的超级块

对于 576i25 DV25 和 DVCAM 的 576i25 变体，选择 4：2：0 色度采样，据称提供了对 DVB 的兼容性。4：2：0 超级块结构如图 46-1 中间的图所示；在 4：2：0 模式下，每个超级块包含一个规则的 9×3 宏块数组。

对于消费类 480i29.97 DV25，选择 4：1：1 亚采样。这足以达到 NTSC 色度带宽和避免亚采样与隔行扫描的冲突。适合于 4：1：1 亚采样的 BT.601 视频每行有 180 个有效色度采样点；但是 180 不能被 8 整除，所以不可能用一个规则的超级块结构。为了形成 8×8 色度块，图像右边沿的 C_B 和 C_R 采样点被特殊处理，详见图 46-2：在垂直方向上相邻的一对由每行最右端色度采样点构成的 4×8 矩阵经水平对接形成 8×8 块。

DV25 的演播室版本（用于 D-7）对 480i29.97 和 576i25 都使用 4：1：1 色度亚采样；使用了特殊的 4：1：1 超级块结构。

DV50 有 DV25 2 倍的数据传输速率。DV50 对 480i29.97 和 576i25 都使用 4：2：2 色度亚采样。由于 4：2：2 亚采样，每个宏块有四个块（不像是 4：1：1 或 4：2：0 的六个块）；每帧有 DV25 2 倍的宏块和 2 倍的超级块。4：2：2 超级块结构如图 46-1 顶部的图所示。现在的 DV50 录像机用两个 DV25 编码器实现，处理图像中交替的 24 行的条带；两个编码器将两个通道的处理结果录制到并行的数据轨上。

46.2 DV 帧/场模式

ISO JPEG 假定处理逐行的源图像：所有图像行按空间顺序处理。在视频处理中，隔行帧的两场可以交织到一起，像逐行一样处理，这里每个 8×8 块包含的行轮流取自相对的两场。假如场间运动被限制，这种方案效果不错。然而，当存在场间运动时，DCT 产生的高垂直频率交流系数将有大的幅度。在给定的数据传输速率下，这（两场合并的编码方案）会生成低的图像质量。如果两场分开编码，可获得更高的图像质量。DV 通过允许压缩器在逐块的基础上选择 8-8-DCT 模式或 2-4-8-DCT 模式来动态适应场间运动的程度。

在 8-8-DCT（“帧”）模式下，相对的两场交织在一起，作为一个 8×8 块进行单次 DCT。

在 2-4-8-DCT（“场”）模式下，来自相似两场的采样点形成两个 8×4 数组。两个 8×4 数组逐个元素的和，以及逐个元素的差值要经过独立的 DCT 变换。求和数组与一个直流项相关联；差值数组只有交流项。和与差的 DCT 数组经过之字形扫描，系数交替进行

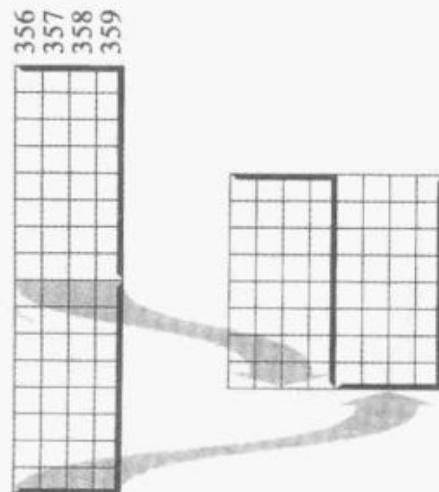


图 46-2 4：1：1 DV 的色度采样

联合 VLE 编码。

此方案包含两个 8×4 DCT，应该称为 2-8-4-DCT！MPEG-2 有一个相似的方案：对于隔行 MPEG-2 序列的每一个宏块，编码器可以选择进行帧 DCT 编码或场 DCT 编码。

DV 标准不强制编码器如何在 8-8-DCT 和 2-4-8-DCT 模式之间进行选择。通常是在空间域里分析亮度差来检测场间运动，如果场间运动显著，就选择 2-4-8-DCT 模式。

46.3 DV 的穿梭图像

VCR 在快进和快退（“穿梭”）时恢复一幅可用图像有现实需求。当数字 VCR 工作在穿梭模式时，单独的同步块被从磁带中读出。DV 使每个同步块对应一个编码的宏块 (CM)，从而使在穿梭模式下恢复有用图像成为可能；由此一整个的宏块也可从一个单独的同步块独立重建。这种方法排除了对 DCT 的直流项预测编码外，直流项直接在每个 CM 中表示。

同步块和编码宏块的对应性似乎要求每个宏块编码为固定位数。在 DV 系统中，每个宏块与一个 77 字节的 CM 相关联，并且一个同步块包含一个 CM。每个 CM 包含每个块的 DCT 的直流项和几个交流项。然而，这并不是全部。

46.4 DV 溢出方案

当图像的一部分缺少细节时，其 DCT 的几个低频交流系数可能有显著的幅值，但几乎所有高频交流项都有很小幅度。当图像的一部分有许多细节时，许多高频交流项都有很大幅度。为了精确再现图像，需要记录这些项。一般来说，细节数量的增加需要增加数据容量。

一幅图像每处都包含细节并不常见；通常一幅图像中的复杂部分在空间上是集中的。一种压缩算法应当能通过按需分配位数来适应图像细节的空间分布。如果细节遍布整幅图像，也许一整帧图像都是树叶，那么很可能甚至相当大的重建误差都不会被察觉到。

穿梭图像的需求似乎不能将数据容量分配给需要更多位的区域。不过，DV 实现了一种溢出方案，根据这种方案，对少量宏块压缩产生的位可以在少量 CM 中共享。假如一个复杂块的 VLE 的交流系数（附有 4 位 EOB）需要比分配给该块的固定容量更多的位，溢出的位会“溅”到其他容量未被填满的块中。溢出的数据首先填入其他具有相同 CM 的块中可能剩余的空间；仍剩余的位填入与图像的不同区域相关联的任何四个其他 CM 中的可用空间。共享溢出空间五个编码宏块的集合称为段 (segment)；一个段有 385 字节的固定容量。一个段中的宏块在空间上分布于整个帧，没有两个宏块来自超级块的同一行或同一列。这种分布利用了统计学的似然性，即只有一个或两个宏块是复杂的。

一个段中的 CM 按从 a 到 e 标记。

当 DV 录像机工作在穿梭模式时，可能单独一个 CM 会独立地被复原，而没有段中的任何其他四个 CM。宏块的溢出数据很可能丢失。不过，溢出数据限于高频交流系数。在穿梭回放时，丢失的溢出系数用零替代。这造成重建图像的细节损失，但是这些系数的缺失不会严重降低图像质量，而且在任何情况下用户也不指望穿梭模式下的图像质量和普通回放模式下的一样。

46.5 DV 量化

DV 编码的主要挑战是为一个段的交流系数确定合适的量化矩阵，以便在所有的量化系数进行可变长编码时，VLE 编码的系数恰好能放进可用的空间。目标是尽可能精细量

化交流项而不超出段的容量。本质上，这是速率控制的一种形式。段的量化发生在 DCT 后，使用以下算法。

首先，段中每一块分为从 0（精）到 3（粗）的等级（class），代表块的空间复杂度。DV 标准提供了一份表格，建议编码器如何根据块中最大的交流项的幅度来分配等级编号；然而，表的使用不是强制性的。

然后，进行多达 15 轮试量化来确定从 0（粗）到 15（精）的量化编号（QNO）。等级编号和量化编号联合起来根据标准中的表格确定量化矩阵。对于每一个测试 QNO，DCT 系数经过量化和之字形扫描。识别出非零交流系数；计算出每一个非零系数对（游程，电平）；累加变长编码的位数。由于量化矩阵的项都是 2 的幂，即（1, 2, 4, 8, 16），量化变得容易；每个系数量化操作只是一次二进制移位。在这一阶段不需要查找和组合 VLE 的位流；眼下只需累计位数。

段最终的 QNO 是产生不超过段容量的最大位数的那个——DV25 的段容量是亮度交流系数 500 位（包括 4 位 EOB）和色度交流系数 340 位（包括 4 位 EOB）。

一旦确定了段的 QNO，就进行 VLE，段中的 CM 组合起来。每个 CM 开始于包含它的 QNO 和误差掩蔽状态（STA）位的一个字节。每个块包含其直流系数、模式（8-8-DCT 或 2-4-8-DCT）和等级。最后，可变长编码的交流系数在一种确定的三轮通过算法中分配：首先是分配到相关的块，然后是分配到同一个 CM 的其他块的未使用空间（如果可利用），最后是分配到段中其他 CM 的未使用空间。选择 QNO 以确保所有的系数有足够的可用空间：每个系数的每位都能存储在段内的某处。

每个 CM 包含 77B，其中包含 1B 头。在 DV25 中，一个 CM 包含 4 个编码亮度块和两个编码色度块。

- 一个总共 14B 的编码的亮度块，包括一个 9 位的直流项，一个模式位，2 位的等级编号。有 100 位可用于交流系数。
- 一个总共 10B 的编码的色度块，包含 9 位直流分量，一个模式位，2 位的等级编号。有 68 位可用于交流系数。

对于 DV50 中的 4:2:2 采样，一个 CM 有四个块而不是六个；在 4:1:1 或 4:2:0 中分配给亮度块的空间可用于溢出数据。对于 3:1:0 采样（用于 SDL，将要讲到），一个 CM 有八个块而不是六个：每个亮度块有 10 字节，每个色度块有 8B。

46.6 DV 数字接口

前文提到的超级块构成了 DV 位流的数字接口的基础。一个 3B 的 ID 置于每个 77 字节的编码的宏块之前，形成 80B 的数字接口（DIF）块。

一个编码的 DV25 超级块用 135 个视频 DIF 块表示，再加上几个非视频 DIF 块形成一个有 150 个 DIF 块的 DIF 序列：

- 1 个头 DIF 块。
- 2 个子编码 DIF 块。
- 3 个 VAUX DIF 块。
- 9 个音频 DIF 块。
- 135 个视频 DIF 块。

实时 DV25 视频需要 10 个或 12 个 DIF 序列，即每秒大约 1500 或 1800 个 DIF 块。DV50 和 DV100 系统有类似的结构，但数据传输速率不同。

在 DIF 序列中打包后，DV 位流可以通过 IEEE 1394 接口传输（也称为 FireWire 和

i. LINK)。IEEE 1394 适用于消费类应用，且广泛应用于桌面视频。对于专业应用，DIF 序列位流可以通过各种接口传输，包括 SDTI 接口（通过 SDTI 传输 DV 时用不到 3 字节的 ID）。DIF 序列可以存储在文件中，例如存储在 QuickTime 或 .mxf 文件中。

46.7 消费类 DV 录制

DV25 广泛用于在 MiniDV 磁带上录制消费类标清内容。几种将 DV 扩展到高清的方案在本书第一版已有介绍；这些方案均没有商业化。在 MiniDV 上录制消费类高清内容用的是 HDV 系统。

46.8 专业 DV 变体

DV 技术为消费类视频磁带录制而推出，并广泛用于消费类摄像机和桌面视频编辑系统。DV 视频磁带技术为适应专业磁带录像机做了修改——首先是用于标清的 DVCPro (D-7) 和 DVCAM，然后是用于高清的 DVCPro HD (D-12)。DV 压缩系统过渡到使用硬盘驱动和闪存媒介的产品。

DVCPro 和 DVCAM 的数据传输速率为 25Mb/s，与消费类 DV 的相同；然而，物理磁带更鲁棒，更适合专业用途。

同样适用于标清的 DV50 有 2 倍于 DV25 的数据传输速率、每秒（每帧）宏块数、每秒（每帧）超级块数。DV50 使用 4：2：2 亚采样；一个 CM 只包含两个亮度块而不是四个；在 DV25 中分配给其他两个块的亮度系数的空间可用于溢出交流系数。对应的直流分量、模式位、等级位等得到保留。在 DV50 中，DV25 交流系数空间的前 4 位被 EOB 符号填充。

DVCPro50 DVTR 被 SMPTE 的 D-7 系列标准化。DV50 标清位流被录制到 6mm DVC 样式的磁带上。

DV100 将 DV50 数据传输速率翻倍达到 100Mb/s，并支持高清图像格式和 4：2：2 色度亚采样：从 1080i 或 1080p 亚采样得到的 1280×1080 像素图像，或从 720p 亚采样得到的 960×720 像素图像。

Panasonic 推出 DVCPro 高清 DVTR 格式，后来标准化为 SMPTE D-12。

第47章

MPEG-2 视频压缩

M-JPEG 核心的基于 DCT 的场内或帧内压缩适合于视频制作；然而，为了内容分发，使用帧间编码可获得高得多的压缩率。MPEG-2 视频压缩利用了时间相关性——视频序列中的连续图像是非常相似的这种统计似然性。MPEG-2 预期的应用从低于标清到超过高清；预期的数据传输速率从约 1.5Mb/s 到大于 20Mb/s。MPEG-2 也定义了音频压缩，并为视频与相关联的音频一起传输做了准备。

MPEG-2 指一组由 ISO/IEC 和 ITU-T 联合颁布的标准。该标准开始于“第一部分：系统”和“第二部分：视频”，由 ISO、IEC 和 ITU-T 联合颁布。另外七个部分由 ISO 和 IEC 联合颁布——“第三部分：音频”、“第四部分：一致性测试”、“第五部分：软件仿真”、“第六部分：DSM-CC 扩展”、“第七部分：先进音频编码（AAC）”、“第九部分：系统解码器的实时接口扩展”和“第十部分：数字存储介质命令与控制（DSM-CC）的一致性扩展”。已立项的第八部分用于 10 位视频，已终止。MPEG-2 标准首次于 1996 年发布，随后又发布了一些勘误与修订。

见 ISO/IEC 13818-1《运动图像和相关音频信息的通用编码：系统》[MPEG-2]，也作为 ITU-T H.220.0 颁布。

ISO/IEC 13818-2《运动图像和相关音频信息的通用编码：视频》[MPEG-2]，也作为 ITU-T H.262 颁布。

MPEG-2 详细规定了是什么构成了一个合法位流：一个合法的（“一致的”）编码器只产生合法的位流；一个合法的解码器可正确解码任何合法的位流。MPEG-2 没有对编码器如何实现压缩进行标准化！

MPEG-2 标准隐含地详细定义了解码器是如何从编码的位流重建图像数据的，而没有强制解码器的实现方式。MPEG-2 明确地避免规定所谓“显示过程”，即重建图像是如何显示的。大多数 MPEG-2 解码器的实现都有灵活的输出格式，然而，MPEG-2 解码设备一般都设计为输出一种特定的光栅标准。

一个 MPEG-2 位流可以表示隔行或逐行的图像。典型的解码器设备输出隔行或逐行的信号。特定的解码设备可以在两种输出格式间切换。因为隔行扫描在标清和高清消费类电子设备中仍占主流，一个解码系统必须能够从逐行序列产生隔行信号。MPEG-2 解码器具有空间重采样能力也是有实际需要的：如果将标清序列输入高清 MPEG-2 解码器，假如重建图像没有经上变换显示为高清，消费者会抱怨的。

47.1 MPEG-2 的类和级

在解码器端，一个 MPEG-2 位流可能会调用许多算法特性，一些从业者称之为“工具”。另外，位流能反映出许多可能的参数值。MPEG-2 标准用一个类（profile）和级（level）的矩阵对位流和解码器进行分类。

MPEG-2 规定了几种算法特性，像是任意帧频和 4:4:4 色度亚采样——这些特性在任何标准类中均不允许。这些特性不大可能商业化。

类限制了编码器、位流、解码器中可能用到的算法特性。类越高，解码器所需的复杂度越高。MPEG-2 定义了 6 种类：简单类 (SP)、主类 (MP)、4:2:2 类 (422P)、信噪比 (SNR)、空间类 (Spt)、高类 (HP)，以及多视点类 (MVP)。

级限制了编码器或解码器使用的参数值。级越高，解码器需要的存储量越大，或是数据吞吐率越高。MPEG-2 定义了 4 种级：低级 (LL)、主级 (ML)、1440 高级 (H14) 和高级 (HL)。

类和级的组合以类和级以@号分隔开的形式表示，例如 MP@ML 或 MP@HL。信噪比、空间、高和多视点这几类与视频的制作和发布无关，不太可能见到商业应用，本书不予进一步讨论。

表 47-1 总结了 MPEG-2 定义的类和级的组合，不包括 SNR、Spt、HP 和 MVP 类。除了 422P，各组合具有层次关系：一个声称符合任何一级的解码器必须能解码所有更低级的视频。作为例外，简单类主级 (SP@ML) 解码器必须能解码主类低级 (MP@LL)。

表 47-1 用不同列表示 MPEG-2 的类（未包括 SNR, Spt, HP 和 MVP），类规定了算法特性；用不同的行表示 MPEG-2 的级，级限制了参数值。表中每一项给出最大画面尺寸、帧频和数据传输速率。两个阴影项在商业上占主导：主类主级 (MP@ML) 用于标清分发，主类高级 (MP@HL) 用于高清分发。SMPTE 308M 对 422P@HL 的图像组结构进行了约束。任何符合 MPEG-2 的解码器必须能解码 MPEG-1 的限定参数位流 (CPB)，受限的参数实际上组成了类/级的组合。

表 47-1 MPEG-2 的类和级

类@级	MPEG-1 CPB	简单 (无 B 图像)	主 (MP)	4:2:2 (422P)
高 (HL)			1920×1152 像素 60Hz 80Mb/s	1920×1088 像素 60Hz 300Mb/s
高-1440 (H14)			1440×1152 像素 60Hz 47Mb/s	
主 (ML)		720×576 像素 30Hz 15Mb/s	720×576 像素 30Hz 15Mb/s	720×608 像素 30Hz 50Mb/s
低 (LL)			352×288 像素 30Hz 4Mb/s	
MPEG-1 CPB *max 99kPx	768×576 [*] 像素 30Hz 1.856Mb/s			

每一个符合 MPEG-2 标准的解码器必须能够解码 MPEG-1 的限制参数位流 (CPB)。表 47-1 还列出了 MPEG-1 CPB，就好像它也由类和级组成，以强调 MPEG-2 一致性要求。

简单类没有 B 图像。禁用 B 图像使编码延迟最小，并使解码器的缓冲存储最小。然而，简单类缺乏 B 图像的压缩效率。

在表 47-1 的八种组合中，只有两种在电视广播中有商业上的重要性。MP@ML 用于标清分发和 DVD，数据传输速率大约 2Mb/s 到 6Mb/s。MP@HL 用于高清分发，数据传输速率通常为 10Mb/s 和 20Mb/s。

4:2:2 类允许 4:2:2 色度亚采样，设计用于电视制作。有一个单独的 4:2:2 类

的主要原因是主类不允许 4:2:2 色度亚采样。MPEG-2 的高类允许 4:2:2 亚采样，但要求高类的一致性会迫使解码器处理信噪比和空间的可分级性。422P@ML 作为 Sony MPEG IMX 用于演播室，数据传输速率为 30~50Mb/s。表 47-2 给出了一些主类和 4:2:2 类的数值化参数限制。

422P@ML 允许在 25Hz 帧频下有 608 行，但在 29.97Hz 和 30Hz 帧频下限制到 512 行。

DVD-视频规范要求符合 MP@ML，并加了额外的约束。720p、1080p 和 1080i 高清的 ATSC 标准要求符合 MP@HL，并加了额外的约束。

表 47-2 MPEG-2 主类和 4:2:2 类

类@级列	图像 (N_c)	图像行 (N_R)	帧频/Hz	亮度采样值速率 [samples/s]	数据传输速率 / [Mb/s]	VBV 容量 /KB
422P@HL	1920	1088	60	62 668 800	300	5 760
MP@HL	1920	1088	60	62 668 800	80	1 194
MP@H-14	1440	1088	60	47 001 600	60	896
422P@ML	720	608	60	11 059 200	50	1 152
MP@ML	720	576	30	10 368 000	15	224
MP@LL	352	288	30	3 041 280	4	58

MPEG-2 定义了 4:2:2 类高级 (422P@HL)。除了 MPEG-2 对 422P@HL 的要求外，SMPTE 308M 规定了对高速率下允许使用的图像组结构限制，如表 47-3 所示。

见 SMPTE 308M。

前面已经给出 MPEG-2 本身的类和级的限制。具体的 MPEG-2 应用，像是 ATSC 广播和 DVD，在 MPEG 的类和级之上还施加了限制。

例如，MPEG-2 允许 24Hz 的帧频，但该帧频不允许用于 DVD (原来是 24 帧/s 的电影一般编码到 480i 的 DVD 时帧频是 29.97Hz，但标记了 2-3 下拉)。

47.2 图像结构

MPEG-2 中每帧都以固定数量的图像列 (S_{AL} , MPEG 中称为水平尺寸) 与图像行 (L_A , 称为垂直尺寸) 的亮度样值进行编码。作者使用亮度这个词；MPEG 文件用的是照度，但在 MPEG 的上下文中照度这个词在技术上是不正确的。

关于照度与亮度的区别，见附录 A。

MPEG-2 的帧要么是正方形采样，要么是有 4:3、16:9 或 2.21:1 的图像幅型比，也就是说，只允许有 4:3、16:9 或 2.21:1 的非正方形采样 (MPEG 的幅型比写法与传统的高度:宽度不一样)。表 47-4 给出 MPEG-2 幅型比信息字段。2.21:1 的值在任何已定义的类中都不允许使用。

MPEG-2 既支持逐行的素材，也支持隔行的素材。一幅有 N_c 列和 N_R 行亮度采样值的图像可直接作为图 47-1 所示的帧结构图像来编码。在帧结构的图像中，1 帧中所有的亮度和色度样值都在同一时间获取，并设计为在同一时间显示。“逐行序列”标志表明序列

表 47-3 SMPTE 308M 中的图像组限制

数据传输速率 (Mb/s)	是否为隔行？	允许的 GoP 结构
0~175	任意	任意
175~230	任意	仅为 I、IP 或 IB
230~300	隔行 逐行	仅为 I 仅为 I、IP 或 IB

表 47-4 MPEG-2 幅型比信息

编 码	幅 型 比
0000	禁止
0001	正方形采样
0010	4:3
0011	16:9
0100	2.21:1
0101…1111	保留

只包含帧结构的图像。

另外，一个视频帧（通常来自隔行的信号源）可以作为一对场结构图像编码——一个顶场图像和一个底场图像，每幅图像都有 N_C 列和 $N_R/2$ 行，如图 47-2 所示。这两个场在时间上差半帧，并且设计为隔行显示。场图片总是成对出现，具有相反的奇偶性（顶场/底场）。一对场中的两幅图像必须有相同的图像编码类型（I、P 或 B），除非 I 场后可跟着一个 P 场（在这种情况下，这一对场作为 I 帧，并且可能用术语 IP 帧表示）。

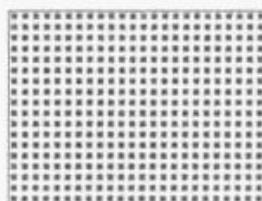


图 47-1 MPEG-2 帧图像

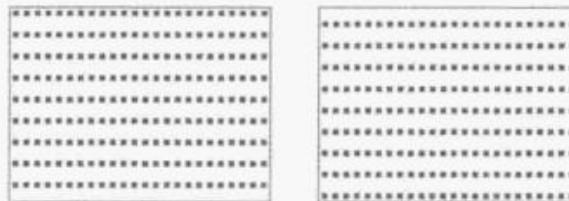


图 47-2 MPEG-2 场图像对

47.3 MPEG 中的帧频和 2-3 下拉

表 47-5 显示了 MPEG-2 的类支持的显示帧频。在同一个视频序列内帧频是恒定的。不幸的是，未规定译码器要用多久来适应帧频的改变。

表 47-5 MPEG-2 帧频的编码

编 码	帧频/Hz	编 码	帧频/Hz
0000	禁止	0101	30
0001	24/1.001	0110	50
0010	24	0111	60/1.001
0011	25	1000	60
0100	30/1.001	1001, …, 1111	保留

在帧结构的图像序列中，为了使显示过程能够实现 2-3 下拉显示而规定了在 MPEG-2 位流中包含相关信息。在这样的一个序列中的帧按帧结构图像编码；在每一帧中，两场都对应同一时刻。图像可能伴随着一个“重复第一场”的标志；如果该标志被置位，则隔行显示器应显示第一场、第二场，然后又是第一场，即进行 2-3 下拉。位流中的帧频指定了 2-3 处理后的显示速率。图 34-1 给出了四幅电影帧的 2-3 序列；在 DVD 上，该序列将按带有表 47-6 给出的标记的一组四幅逐行扫描的 MPEG-2 帧来编码。

帧编码的图像可对一个顶场/底场对或底场/顶场对进行编码，即帧图像可能对应一个视频帧，或是把两个视频帧叠在一起。后一种情况支持 2-3 下拉中的 M 帧。

表 47-6 MPEG-2 中 2-3 下拉序列

电 影 帧	顶上第一场 (TFF)	重复第一场 (RFF)
A	0	0
B	0	1
C	1	0
D	1	1

47.4 亮度和色度采样结构

MPEG-2 支持适合演播室应用的 4:2:2 色度亚采样和适合视频分发的 4:2:0 色度

亚采样。不同于 DV, C_B 和 C_R 样值对在空间上是重叠的。MPEG-2 标准包括 4:4:4 色度格式，但在任何已定义的类中都是不允许的，所以极不可能被商业化。

4:2:2 没有垂直亚采样，此时亚采样和隔行扫描互不影响。图 12-1 所示的第三列 (BT. 601) 为帧结构和场结构图像的 4:2:2 色度。

帧结构图像的 4:2:0 色度亚采样如图 12-1 最右边一列所示； C_B 和 C_R 采样点在垂直方向上位于帧中的亮度采样点的正中，在水平方向上位置相同。

图 47-3 给出了场结构图像中的 4:2:0 色度亚采样。在顶场中，色度采样点在垂直方向上位于同一场中的上下两个亮度采样点的 1/4 处（在本例中， C_{B0-3} 位于从 Y'_0 向下到 Y'_2 的 1/4 处）。在底场中，色度采样点在垂直方向上位于同一场中的上下两个亮度采样点的 3/4 处（在本例中， C_{B4-7} 位于从 Y'_4 向下到 Y'_6 的 3/4 处）。该方案把色度采样点放在与帧结构图像的色度采样点相同的位置上，只不过在帧中交替的色度采样行有半帧的时间偏移。

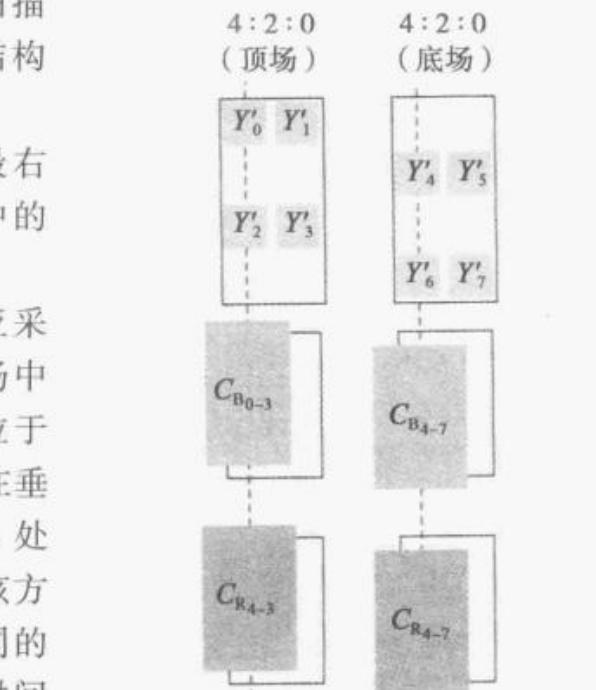


图 47-3 场结构图像中的色度亚采样

47.5 宏块

MPEG 压缩的核心是对采样值的 8×8 块（在 I 图像中和 JPEG 一样），或对预测误差（残差，在 P 和 B 图像中） 8×8 块的离散余弦变换 (DCT) 编码。为了简化色度亚采样的实现，对亮度和色度使用同样的 DCT 和块编码方案。当结合 4:2:0 色度亚采样时，两个 8×8 色度采样块与一个 16×16 亮度块相关联。这使得场或帧由 16×16 亮度采样值的单元拼接而成。每一个这样的单元称为宏块 (MB)。宏块位于与图像左上角的亮度采样值对齐的 16×16 的网格上。

如果水平尺寸或垂直尺寸不能被 16 整除，则编码器会在图像的右边缘或下边缘填充合适数量的黑色“延伸”样值。例如，对于 1920×1080 像素高清编码，编码器在图像阵列后添加 8 行的黑色像素，使行数为 1088。在解码时，这些样值在显示之前被切掉。延伸的区域保存在参考帧存储区。

每个宏块包括四个 8×8 亮度块，以及随色度格式而定的规定数量和排列的 $8 \times 8 C_B$ 块和 $8 \times 8 C_R$ 块。在常见的 4:2:0 色度格式中，一个宏块包含六个块：四个亮度块、一个 C_B 块和一个 C_R 块。在 4:2:2 色度格式中，一个宏块包含八个块：四个亮度块、两个 C_B 块和两个 C_R 块。

47.6 图像编码类型 (I、P、B)

16.8 节介绍了 I 图像、P 图像和 B 图像。编码的 I 图像和 P 图像数据用于重建参考图像，也就是可用于构成预测的场或帧。MPEG 解码器保持两个参考帧存储，一个过去帧和一个未来帧。编码器也保持两个当作是由解码器重建的参考帧存储；编码器的参考帧存储跟踪着解码器参考帧存储的内容。简单类没有 B 图像，一个参考帧存储就够了。

MPEG-2 和 H.264 使用术语“参考图像”。有人称其为“锚图像”。

每一幅 I 图像的编码与任何其他图像无关。当解码器重建出 I 图像时，I 图像被显示。

另外，它被存储为一个参考帧，可作为预测器使用。I 图像通过类似 JPEG 的算法压缩，使用基于感知的量化矩阵。

每一幅 P 图像用过去的参考图像作为预测器进行编码。残差用与 I 图像同样的类似 JPEG 算法进行压缩，不过通常使用的量化矩阵大不相同。当解码器重建出 P 图像时，P 图像被显示；另外，P 图像被写入到参考帧中，以用于后续预测。

每一幅 B 图像包含根据一个或两个参考帧双向预测的元素。编码器计算，压缩并传输残差。解码器重建 B 图像并显示，然后丢弃之：B 图像都不用于预测。

每一幅参考图像与一整帧的存储相关联。当解码器重建参考场（I 场或 P 场）时，参考帧存储的一半行被写入，另一半保留先前的参考场的内容。场对的第一场被重建后，就能够作为第二场的预测器（先前参考帧的第一场不再有效）。

47.7 预测

图 16-1 给出了一个简单的图像间编码方案。对于从一个视频帧到下一个帧移动超过几个像素的任何场景元素，简单方案会产生大量的图像间差值。编码器进行运动补偿预测可以更有效地对运动进行编码。编码器还生成运动向量；运动向量用于置换参考帧中的一部分区域，相对于非置换的预测，能提高对当前图像的预测效果。残差再用 DCT、量化、VLE 进行压缩。

在解码器端，根据传输的运动向量和预测模式，由参考图像形成预测。通过 VLE 解码、反量化和反 DCT 从位流恢复出残差。最后，解码后的残差加到预测值上形成重建图像。如果解码器正在重建 I 图像或 P 图像，重建的图像被写入到参考帧的合适部分（或是全部）。

编码器形成前向图像间差值的显而易见方法是，要从参考图像减去当前的源图像（参考图像可能要根据编码器的运动估计进行运动补偿内插）。从帧内编码图像开始，解码器便会累积图像间的差值。然而，MPEG 包含有损压缩：图像组的起点 I 图像和每一组解出的图像间差值都会有重建误差。用计算图像间差值的简单方案，重建误差会在解码器端累积。为了缓解这种解码器误差的潜在来源，编码器集成了解码器。图像间差值是通过从解码器将重建出来的先前的参考图像中减去当前的源图像而得到。重建误差由此变成只在环路内部，而且不会累积。

MPEG-2 使用的预测模型是亮度采样值的 16×16 块（与相关联的色度采样值一起）的逐块平移：当前图像的宏块是从重建的参考图像中一个类似尺寸的区域预测的。 16×16 的区域尺寸选择是对使用大的区域（有效地利用空间一致性，以及把运动向量的开销分摊到相当大数量的采样值上）和小区域（有效地编码运动的小场景元素）要求的折中。

参考帧中的预测区域很少与 16 个亮度采样点的宏块网格对齐；称其为宏块是不恰当的。有的作者没能区分宏块和预测区域；别的作者用术语“预测宏块”表示预测区域。

P 图像的宏块一般是经前向预测得到的。不过，编码器可以决定一个特定宏块最好采用帧内编码（即，完全不预测）。B 图像的宏块通常被预测为经运动补偿的过去和未来参考图像的均值，即通常是双向预测的。不过，编码器可以决定 B 图像的一个特定宏块最好采用帧内编码，或采用前向或后向的单向预测。表 47-7 给出四种宏块类型。在任何图像中允许使用的宏块类型受图像类型的限制，如表 47-8 所示。

在一个闭合的图像组中，B 图像不能做下一个开始图像组的 I 图像的前向预测图像。

表 47-7 MPEG 宏块类型

		预 测	典型的量化矩阵
帧内		无——宏块是自我包含的	基于感知的
帧间（非帧内）	后向预测编码	根据将来的参考图像预测	平坦
	前向预测编码	根据过去的参考图像预测	平坦
	双向预测编码	根据过去和将来的参考图像取均值预测	平坦

表 47-8 MPEG 图像编码类型

	二进制代码	是否有参考图像	允许的宏块类型
I 图像	001	是	帧内
P 图像	010	是	帧内 前向预测编码
B 图像	011	否	帧内 前向预测编码 后向预测编码 帧内 双向预测编码

隔行扫描序列中的每个非帧内宏块要么通过帧预测（一般当场间运动很小时，编码器选择该模式），要么通过场预测（一般当场间运动显著时，编码器选择该模式）来进行预测。这类似于前文介绍的 DV 中的场/帧编码。场图像的预测器必须是场预测器。然而，帧图像的预测器可以逐个宏块地选择场预测器或帧预测器。MPEG-2 定义了几个可以在逐个宏块的基础上选择的额外的预测模式。表 47-9 总结了 MPEG-2 的预测模式。

表 47-9 MPEG-2 预测模式

	用 于	描 述	运动向量最大个数	
			后 向	前 向
帧预测	P 图像、B 图像	对帧进行预测，使用之前重建的一或两帧数据	1	1
场预测	P 图像、B 图像 P 场、B 场	对每一场独立进行预测，使用之前重建的一或两场数据	1	1
16×8 运动补偿 (16×8MC)	P 场、B 场	分别对上方 16×8 区域和下方 16×8 区域进行预测（与顶场、底场完全无关）	2	2
双主场	没有插进 B 图像的 P 场	从传输的运动向量和一个小的运动向量差值得到两个运动向量 (DMV, -1, 0 或 +1)，用于从两个参考场（一个顶场，一个底场）形成预测值，并求均值得到预测器	1	1
双主帧	没有插进 B 图像的 P 图像	与上面的方法类似，但是需要对 2 场重复操作，得到 4 个预测值并求均值	1	1

47.8 运动向量

运动向量 (MV) 确定了参考图像中用于预测的一块 16×16 亮度采样值区域。运动向量指向可能在空间上离待编码区域相当远的一块预测区域，也就是说，运动向量的范围可以相当大。即使在场图像中，运动向量也以帧的亮度采样点为单位来指定。运动向量可以指定为整像素坐标，在这种情况下仅复制像素就可完成 16×16 预测。然而，在 MPEG

中，运动向量可以被指定为半像素精度：如果运动向量的小数位被设置，则对邻近的整数坐标上的采样值求平均（即线性内插）形成预测值。传输的运动向量值减半用于亚采样的色度。所有定义的类要求不能有运动向量指向参考图像边界外的任何采样点。

每个宏块的头部包含运动向量的计数。运动向量本身也是要预测的！初始的运动向量建立在宏块条的开头；每一个后续非帧内宏块的运动向量参考按光栅扫描顺序的前一个宏块进行差分编码。

运动向量进行可变长编码，以便于短的向量——在平移运动或不运动的大片区域内最可能出现的运动向量——进行紧凑编码。零值运动向量非常可能出现，所以为其紧凑编码做了规定。

帧内宏块不经过预测，所以不需要运动向量。然而，在某些情况下允许有掩错运动向量（CMV）：如果由于传输错误丢失了一个宏块，CMV 允许解码器使用其预测机制来合成图像信息，以掩盖错误宏块。CMV 如果包含在它自己的宏块中将没有用处。因此，CMV 是与紧下方的宏块相关联的。

47.9 块编码

每个宏块伴随有少量的预测模式信息：零个、一个或多个 MV 和 DCT 编码的残差。

帧内宏块中每个块的编码与 JPEG 的块类似。变换系数用一个（正常情况下）根据感知特性加权的量化矩阵进行量化。规定了 8 位、9 位、10 位的直流系数（在 422 类 [422P] 中，允许 11 位的直流系数）。在一个宏块条内，对直流系数进行差分编码。

在 I 图像中，DCT 的直流项进行差分编码：每个亮度块的直流分量用作下一个宏块的对应直流分量的预测器。 C_B 和 C_R 块的直流分量进行类似的预测。

原则上，非帧内宏块的残差可以直接编码。在 MPEG 中，出于两个原因，残差用 DCT 编码。首先，DCT 编码利用残差中可能出现的空间一致性。其次，DCT 编码允许使用已用于帧内宏块同样的（基于量化的）码率控制以及 VLE 编码。非帧内块的残差进行反量化，然后添加到从参考帧经运动补偿得到的值上。因为反量化的变换系数不能直接观察到，所以不适于使用根据感知特性加权的量化矩阵。默认情况下，非帧内块的量化矩阵是平坦的，即所有项有相同的值。

47.10 帧和场 DCT 类型

宏块中的亮度按照帧 DCT 编码或场 DCT 编码这两个方案之一划分为四个块。下面先介绍适于帧 DCT 编码的三种情况，然后介绍场 DCT 编码。

- 从一个原生的逐行源得到的帧结构图像中，每一个宏块用参考帧中在空间上连续的 16×16 区域做预测效果最好。这是帧 DCT 编码：宏块的亮度采样值划分成 8×8 的亮度块，如图 47-4 所示。

帧 DCT 类型将每个 16×16 宏块的亮度采样值分割为四个 8×8 块。这对场图像、原生的逐行帧图像和场间运动很小的帧结构图像中的宏块最有效。

- 在场结构图像中，源帧的交替的图像行都被编码器拆到两个场中，每一行都没有隔行效应。在这样一幅图像中的每一个宏块用参考帧中在空间上连续的 16×16 区域（或理解为参考帧的 16×32 区域的交替行）做预测效果最好。这也是帧 DCT 编码。
- 在来自隔行源的帧结构图像中，不含有运动的场景元素的宏块通常用帧 DCT 编码预测最好。

在来自隔行源的帧结构图像中，当宏块含有运动的场景元素时，需要使用另一种方

案。这样的场景元素在第一场和第二场中的位置不同，因此参考图像中一个空间连续的 16×16 区域会形成比较差的预测器。MPEG-2 提供了一种有效地编码此类宏块的方法。该方案将亮度交替分割成 8×8 的块：通过取参考帧中的交替行形成亮度块。该方案称为场 DCT 编码，如图 47-5 所示。

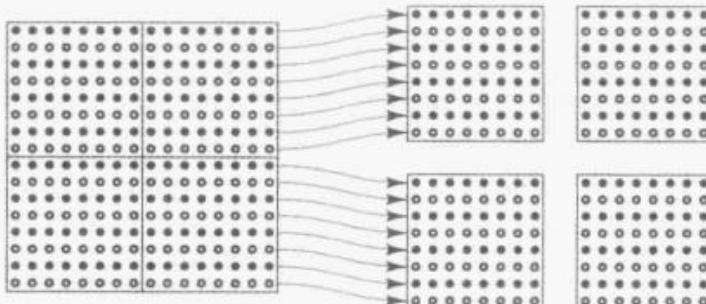


图 47-4 帧 DCT 类型

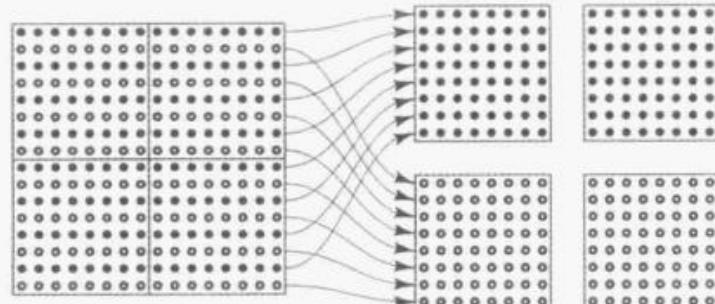


图 47-5 场 DCT 类型

场 DCT 类型通过取交替的图像行产生了四个 8×8 亮度块。这允许对有显著场间运动的来自隔行源的帧结构图像进行有效编码（在场结构的图像中已经隐含了类似的解交织）。

读者可能认为用和处理亮度采样值相同的方法来处理隔行帧图像的色度采样值是一个好主意。然而，经过 $4:2:0$ 亚采样，同样的处理方法就需要有 8×4 色度块或 16×32 宏块。这两种选择都不是希望的。因此，在有场间运动的帧结构图像中，色度块一般预测效果差劲。由于 $4:2:2$ 色度格式中没有垂直亚采样， $4:2:2$ 的序列从本质上不会有糟糕的色度预测。

47.11 之字形扫描和可变长编码

一旦 DCT 系数被量化，编码器以之字形顺序扫描这些系数。JPEG 的之字形扫描如图 47-6 所示。图 47-6 所示的扫描顺序也用于 MPEG-1。

除了 JPEG/MPEG-1 扫描顺序外，MPEG-2 提供了另一种为来自隔行源帧结构图像优化过的扫描顺序。这种替代的扫描方案，如图 47-7 所示，可以在逐幅图像的基础上由编码器进行选择。

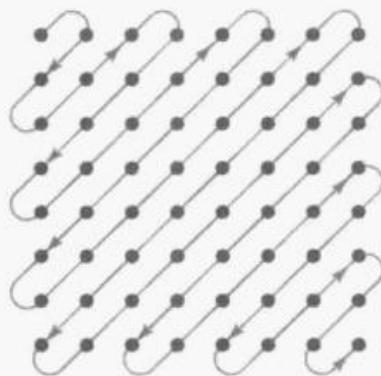


图 47-6 之字形扫描 (1)

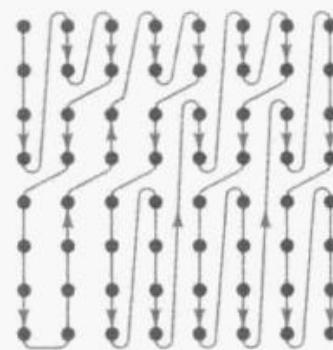


图 47-7 之字形扫描 (2)

之字形扫描后，确定了零值交流系数，然后形成〈游程，电平〉对，并进行可变长编码。对于帧内宏块，MPEG-2 允许编码器在两种 VLE 方案之间选择：一种是首先在 MPEG-1 中标准化的方案，另一种方案更适合于有场间运动的帧结构图像。

MPEG-2 编码器和解码器系统的框图如图 47-8 所示。编码器包含了一个运动估计器，涉及巨大的计算复杂度。MV 集成到位流中传输给解码器；解码器不需要估计运动。编码

器实际上包含了解码器的一个副本；编码器的图像差值计算的是解码器端可获得的重建图像信息。

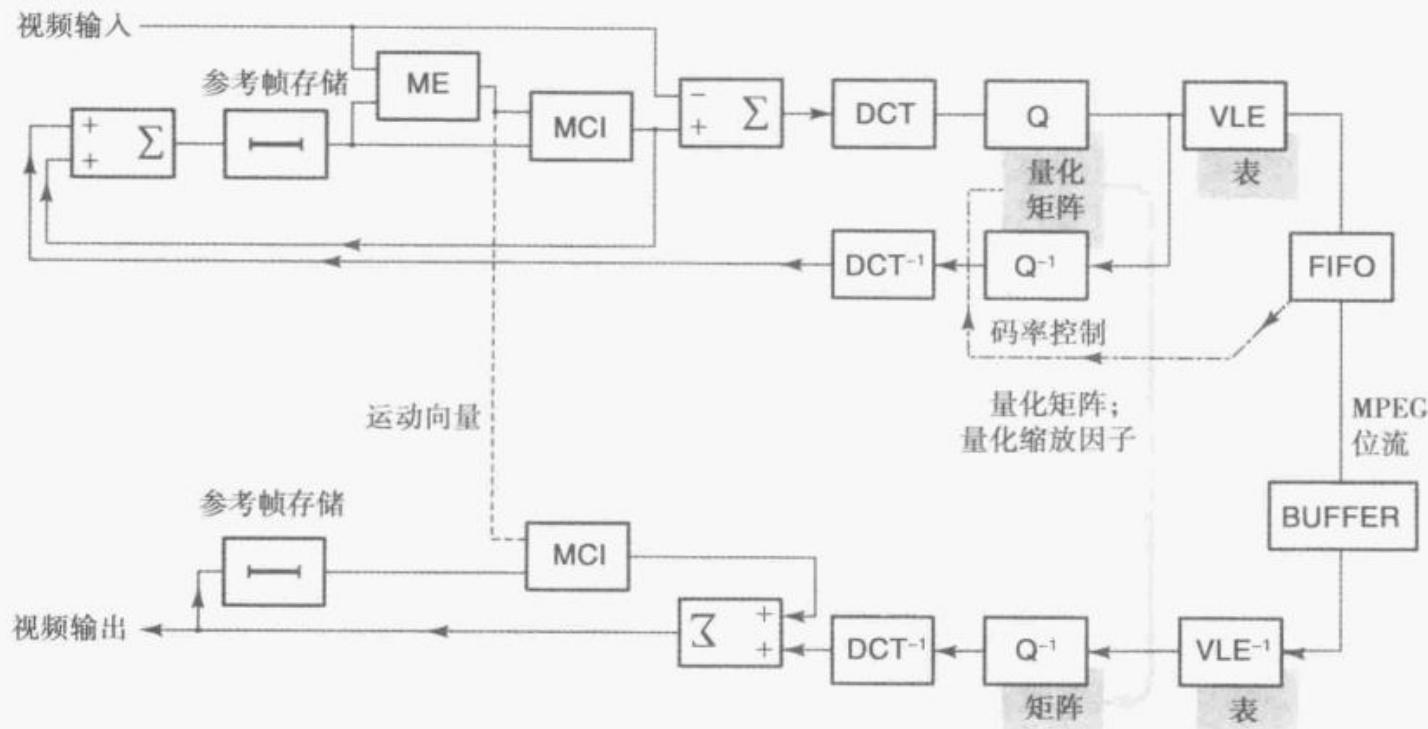


图 47-8 MPEG-2 的编码器和解码器框图

47.12 刷新

出于三种原因，需要时不时地插入 I 宏块：在信道获取时建立一幅参考图像；限制由无法纠正的传输错误引起的伪影持续时间；限制漂移（即由编码器的 IDCT 和解码器的 IDCT 之间没有保持一致而引起的编码器和解码器预测器的差异）。MPEG-2 要求帧中每一个宏块在第 132 个 P 宏块之前用一个帧内宏块刷新。一般编码器通过周期性或间歇地插入 I 图像来满足这一要求。可是，MPEG-2 没有严格要求使用 I 图像，有时候使用的是分散式刷新——I 宏块而不是 I 图像用于刷新，尤其对于直播卫星（DBS）。

一个复杂的编码器检查源视频来检测场景剪切，并根据图像的内容调整图像类型顺序。

47.13 运动估计

运动向量必须比表示从一帧到下一帧的运动做得更多：对于 B 图像，运动向量必须描述从一个参考帧到下一帧的运动，即从一个 I 图像或 P 图像到下一个 I 图像或 P 图像。随着插入的 B 图像数量的增加，例如 M 值增加，运动向量范围也要增加。运动估计的开销和复杂度也随着搜索范围的增加而急剧增加。

运动估计 (ME) 的负担落到编码器头上。运动估计非常复杂而且是计算密集型的。MPEG-2 支持很大的运动向量范围：对于 MP@ML 帧结构图像， 16×16 预测区域可能位于当前被解码宏块的水平方向 $[-1024, \dots, +1023 \frac{1}{2}]$ 和垂直方向 $[-128, \dots, +127 \frac{1}{2}]$ 个亮度采样点之内的任意区域。图像头部的元素 (f 码) 指定每幅图像中使用的运动向量范围；这限制了需要分配给该图像的运动向量的位数。

MPEG 中运动估计的目的不完全是估计图像区域的运动，不如说是为了访问一个使需要被编码的预测误差（残差）信息数量最小的预测区域。一般通过使用 16×16 宏块中平

均运动的最佳估计可达成这一目标，但不总是这样。做出这种区分是因为一些视频处理算法需要精确的运动向量，使估计的运动和人类观察者感知到的运动很好地匹配。很多视频处理算法，像是用于标准转换器的时域重采样，或是去隔行，每个或每几个亮度采样点就需要一个运动向量。在 MPEG 中，只需要一个或两个向量来从一个或两个参考图像的 16×16 区域中预测一个宏块。

如果一个运动向量的分数位被置位，则通过对来自邻近的整数坐标上的像素采样值求平均形成预测值。这对解码器来说很直接。然而，对于编码器来说，要生成水平和垂直方向上的 $1/2$ 亮度采样点的运动向量所需计算工作量是生成整采样点向量的 4 倍。

有三种主要的运动估计方法。

- **块匹配**，也叫做全搜索，涉及从参考帧一定的二维范围内对目标宏块的最佳匹配进行穷举搜索。对于 MPEG-2 的广大搜索范围来说，完全块匹配难以实现。
- **像素递归**，这方法根据先前帧的运动估计从少量的运动初始猜测开始，搜索参考帧中对应的坐标，并细化每个猜测。最好的猜测用作最终的运动向量。
- **金字塔形**，这方法构建目标宏块和参考帧的空间低通滤波的版本，在低分辨率下进行块匹配。在一个分辨率级别上，围绕着最有希望的候选位置坐标，形成参考图像区域的轻度滤波的版本，并在新版本的参考图像区域上进行块匹配。连续细化产生最终的运动向量。这一技术往往会产生平滑的运动向量域。

47.14 码率控制和缓冲区管理

由典型 MPEG-2 编码器编码得到的典型视频序列，产生 I 图像、P 图像和 B 图像消耗位数的大致比例是 60 : 30 : 10。一幅 I 图像可能需要两幅 B 图像 6 倍的位数。

MPEG-2 的许多应用，如数字电视，涉及有固定数据传输速率的传输信道。这需要恒定数据传输速率 (CBR) 操作。其他 MPEG-2 应用，如 DVD，涉及一个有可变（但受限的）数据传输速率的信道。这类应用需要可变数据传输速率 (VBR) 操作，其中瞬时数据传输速率为达到每一帧预期的画面质量，最大限度地利用存储空间而变化。

解码器的缓冲区越大，编码器在图像间分配位就越灵活。然而，大的缓冲区是昂贵的。每个类/级组合规定了解码器必须实现的最小缓冲区的尺寸。

编码器通过改变量化矩阵来影响码率控制——基于感知的加权矩阵用于帧内宏块，平坦矩阵用于非帧内宏块。MPEG-2 允许量化矩阵包含在位流中。另外，更重要的是，量化比例编码在宏块条级传输，并可在宏块级更新。该代码确定用于量化矩阵的总体比例系数。为达到所需的数据传输速率，编码器量化程度或高或低；量化比例编码传送到解码器以使其进行相应的反量化。

视频显示需要每秒有恒定数量的帧。由于 I 图像的位数相对很大，在除了图像质量退化之外的所有情况下，在 I 图像的解码和显示期间，解码器的净缓冲区占用率降低。在 B 图像的解码和显示期间，净缓冲区占用率增加。图 47-9 显示了视频序列开始时约一个图像组持续时间内的典型缓冲区占用率情况。

MPEG 码流的构建必须能使解码器缓冲区不上溢：如果上溢，位会丢失。位流

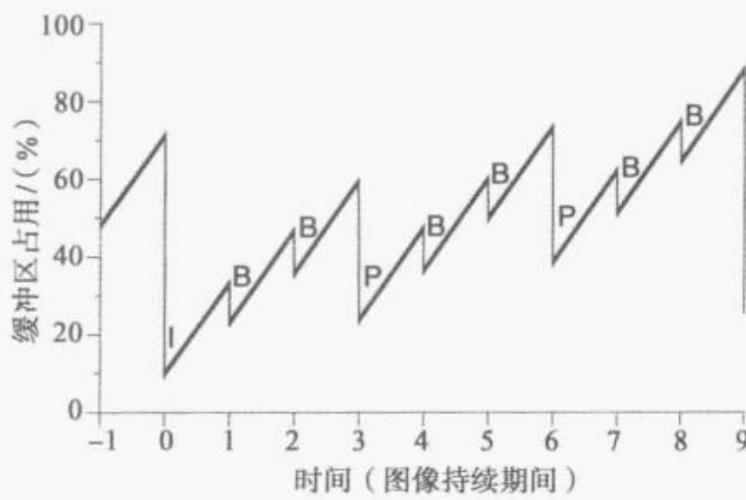


图 47-9 MPEG-2 的缓冲区占用率

的构建必须还能使缓冲区不下溢：如果下溢，待显示的图像就不会在需要的时间可用。

MPEG-2 的缓冲区管理是基于一个理想化的解码器缓冲区模型进行的：与每幅图像相关的所有位被认为在相对于位流的一个精确定义时刻，从解码器缓冲区中提取出来。每个编码器实现一个跟踪理想缓冲区状态的视频缓冲验证器（VBV）。每个图像头部包含一个 VBV 延迟字段，在图像开始时声明缓冲区的充满程度。在信道获取之后，解码器在开始解码前先等待相应时间（如果解码器没有等待，会导致缓冲区下溢）。

47.15 位流句法

MPEG-2 视频压缩的最终结果是一个划分到 MPEG 句法层级中的位流，包含六层：序列（sequence）、图像组（GoP）、图像（picture）、宏块条（slice）、宏块（macroblock）和块（block）。除了视频序列层有序列结束元素外，每个句法元素均有头部没尾部。序列、图像组、图像和宏块条元素均以一个 24 位的开始码前缀开始，其由 23 个 0 位后接一个 1 位组成。开始码建立字节对齐，可能之前有任意数量的零填充位。所有其他数据流元素的构建要避免出现 23 个或更多连零位。

MPEG 句法的顶层是视频序列。序列头部指定高级参数，如数据传输速率、图像率、图像尺寸、图像幅型比、类、级、逐行/隔行扫描以及色度格式。VBV 缓冲区尺寸参数声明序列之内要求的最大缓冲区尺寸。序列头部可能指定量化矩阵。根据编码器的自行决定，序列头部可能会在整个序列中间歇性或周期性地重传，以便于解码器进行快速信道获取。

每个隔行扫描视频序列的开头建立起一个场对的不变序列，次序要么是 480i 典型的〈顶场，底场，…〉，要么是 576i 和 1080i 典型的〈底场，顶场，…〉。在序列之内，任何单独的场可能被场编码，任何两个相邻的场可能被帧编码，但是场的奇偶性必须按严格的顺序轮换。

图像组（GoP）是 MPEG 的随机访问单位。GoP 层在 MPEG-2 中是可选的；然而，GoP 层对于大多数应用都是实际必需的。一个 GoP 开始于一个 I 图像（额外的 I 图像是允许的）。GoP 头包含 SMPTE 时间码、GoP 闭合标志和连接断开标志。

GoP 头部包含 23 位编码的 SMPTE 时间码。如果出现，适用于 GoP 的第一帧（按显示顺序）。在 MPEG 之内不使用。

如果 GoP 是闭合的，GoP 中编码的 B 图像不会参考下一个 GoP 的第一个 I 图像。这样效率不高，因为下一个 I 图像通常包含有用的预测信息。如果 GoP 是敞开的或 GoP 头部未出现，那么 GoP 的 B 图像可参考下一个 GoP 的第一个 I 图像。为了允许编辑 MPEG 位流，GOP 必须是闭合的。

在 GoP 边界拼接位流的设备可以设置连接断开标志，指示解码器立即将 GoP 的第一个 I 图像后的 B 图像置为无效。

图像头部指定图像结构（帧、顶场或底场）和图像编码类型（I、P 或 B）。图像头部可以指定量化矩阵和量化比例类型。VBV 延迟参数用于缓冲区管理。

宏块条层按从左至右从上至下的光栅扫描顺序将宏块集合起来。没有宏块条跨过图像边沿。所有已定义的类都有“受限制的宏块条结构”，宏块条覆盖整幅图像，没有空隙或重叠。宏块条头部包含了量化比例码。宏块条有多种用途。首先，当出现无法纠正的数据传输错误时，宏块条是再同步的最小单元。第二，宏块条是帧内宏块直流分量差分编码的单元。第三，宏块条是非帧内运动向量差分编码的单元：宏块条第一个宏块的运动向量被绝对编码，后续宏块的运动向量以从此之后连续的差值来编码。

宏块是 MPEG 的运动预测单位。编码的宏块数据包含宏块类型的指示（帧内、前向预测、后向预测，或双向预测），量化比例码，0、1 或 2 个前向运动向量，0、1 或 2 个后向运动向量。编码块模式标志提供了一种紧凑的方式表示未编码的块（由于预测不充分，不需要表示残差）。

每个块在位流中表示为 VLE 编码的 DCT 系数——一个差分编码的直流系数，0 或更多交流系数。每个编码块的数据用一个 4 位的块结束（EOB）标志终止。

47.16 传输

MPEG 视频或音频的各种句法元素串行化形成一个基本流（Elementary Stream, ES）。MPEG-2 定义了一种机制将 ES 分成包，形成一个打包的基本流（Packetized Elementary Stream, PES）。每一个 PES 包的头部包含系统级时钟信息、包优先级、包序列号和（可选的）加密信息。如果一个 MPEG-2 PES 存储在文件中，该文件习惯上扩展名为 .m2v；不过，视频 PES 不能包含音频！然而更常见的是 MPEG-2 的视频和音频进行复用，然后用节目流或传送流进行储存或传输，见第 50 章。

延伸阅读

Gibson, Jerry D. Toby Berger, Tom Lookabaugh, David Lindbergh, and Richard L. Baker (1998), *Digital Compression for Multimedia* (San Francisco: Morgan Kaufmann) . Lookabaugh's chapter provides an excellent 55-page description of MPEG-2. His chapter also covers MPEG audio.

Haskell, Barry G. (Atul Puri, and Arun N. Netravali (1997), *Digital Video: An Introduction to MPEG-2* (New York: Chapman & Hall) . This book fails to distinguish luminance and luma; both are called *luminance* and given the symbol Y. See Appendix A, *YUV and luminance considered harmful*, on page 567 of the present book.

Mitchell, Joan L., William B. Pennebaker, Chad E. Fogg, and Didier J. LeGall (1997), *MPEG Video Compression Standard* (New York: Chapman & Hall) . This book concentrates on MPEG-1. Egregiously incorrect information appears concerning chroma subsampling.

Symes, Peter (2003), *Digital Video Compression* (New York: McGraw-Hill) .

Watkinson, John (1999), *MPEG-2* (Oxford: Focal Press) .

H. 264 视频压缩

H. 264 指代由 ITU-T (命名为 H. 264) 和 ISO/IEC (命名为 MPEG-4 第 10 部分) 标准化的编解码器。MPEG-4 第 2 部分的简单演播室类 (SStP) 用于 HDCAM。第 2 部分的这方面内容和第 10 部分的全部适用于广播质量的视频；除了以上情况，MPEG-4 一般不适用于广播质量的视频。H. 264 由联合视频工作组 (JVT) 开发，称为高级视频编码 (AVC)；在开发过程中的 ITU-T 命名是 H. 26L。所有这些术语都曾用于指代当标准被采纳后现名为 H. 264 的标准。

ITU-T H. 264 也作为 ISO/IEC 14496-10 (MPEG-4 第 10 部分) 颁布。

H. 264 与 MPEG-2 大体相似，但摘取了很多“低处的果实”（译者注：即采用了许多现有的简单技术）。H. 264 压缩性能的提升是通过十几项技术获得的，每项也许在编码效率上有 6% 的改善，但十几项级联起来产生了 2 倍于 MPEG-2 的效率（从业者声称 H. 264 效率最低为 MPEG-2 的 1.5 倍，最高为 3 倍）。H. 264 的应用范围广泛，从监控视频，到视频会议，到移动设备，到网络视频流，到 HDTV 广播。

H. 264 是复杂的。标准 (2010-03 版) 包含 669 页详尽描述。实现编码器或解码器需要许多人年。软件、固件和硬件可通过商业途径得到。硬件实现都需要嵌入固件：H. 264 的 VLSI 解决方案通常包含一个或多个嵌入式 RISC 处理器和不少的相关固件。

H. 264 相对 MPEG-2 增加的特性在本章余下部分描述。假定读者熟悉前文的“视频压缩简介”，以及前面三章介绍的 JPEG、M-JPEG、DV 和 MPEG-2。

就像 MPEG-2，H. 264 明确规定了怎样构成一个符合规范的位流：具有一致性（“合法”）的编码器只产生合法的位流；合法的解码器可以正确解码任何合法的位流。H. 264 实际上规定了解码器的行为，但未对编码器进行标准化！

压缩的目的是降低数据速率，同时最小化伪影的可见度。多数专家称评价编码器性能的最好也是唯一的方法就是，对视频流压缩后再解压缩的结果进行视觉评价。

H. 264 受数百项专利保护。标准实现者、制造商、用户和/或其他人会或不会要求从 MPEG LA 管理的“专利池”中取得许可。

并非所有的 H. 264 特性都要在每个解码器中实现；例如，B 宏块条（相当于 MPEG-2 B 图像）在基础类中禁用。各种应用有不同的数据速率，而且解码器可以有不同等级的资源（例如存储器）；就像 MPEG-2，一个类和级的系统确定了最低要求。

48.1 算法特性、类和级

表 48-1 总结了 H. 264 超出 MPEG-2 的算法特性。最上面一段的特性是在所有类中都可用的；下面段中的特性依赖于具体的类。

在基本类和扩展类中可用的特性涉及对通过不可靠信道传输的数据鲁棒性处理。这些特性（和类）一般不关心专业视频，也不允许用于主类和高类。

扩展类、主类和高类的特性提高了编码效率。CABAC 改进了可变长熵编码的性能。

表 48-1 H.264 特性

	算法特征 (“工具”)	类	基准 (BP)	扩展 (XP)	主 (MP)	高 (HiP)
所有类都具备的特征	多参考图像		·	·	·	·
	灵活的运动补偿		·	·	·	·
	I宏块条与 P宏块条		·	·	·	·
	1/4 像素运动补偿内插		·	·	·	·
	16 位精确匹配的整数变换		·	·	·	·
	统一的可变长编码		·	·	·	·
	(UVLC/Exp-Golomb)		·	·	·	·
	CAVLC		·	·	·	·
集合 1	环路内去块滤波器		·	·	·	·
	灵活的宏块排序 (FMO)		·	·		
	任意宏块条顺序 (ASO)		·	·		
集合 2	冗余宏块条 (RS)		·	·		
	数据划分			·		
集合 3	SI 和 SP 宏块条			·		
	B宏块条			·	·	·
	隔行编码 (PicAFF, MBAFF)			·	·	·
集合 4	加权和偏置 MC 预测			·	·	·
	CABAC 熵编码				·	·
FRExt	8×8 亮度帧内预测					·
	加大的样值深度					·
	4:4:4 和 4:2:2 色度亚采样					·
	图像间无损编码					·
	8×8/4×4 变换自适应性					·
	量化缩放矩阵					·
	单独的 C _B 和 C _R 量化步长控制					·
	单色图像 (4:0:0)					·

保真度范围扩展 (FRExt) 是指一些包含在 HiP、Hi10P、Hi422P 和 Hi444P 高类的算法特性，可实现更高质量的视频。HiP 允许 10 位视频；Hi422P 允许 4:2:2 色度亚采样，而 Hi444P 允许 4:4:4、12 位视频和一些别的特性。

四种 H.264 类具有商业重要性：基本类、扩展类、主类和高类。主类和高类与专业视频相关。H.264 有十五个级别，包含从 176×144（编码速率低至 64kb/s）到 4k×2k（编码速率高达 240Mb/s）。对专业视频来说很重要的类与级的组合总结于表 48-2 中。

表 48-2 H.264 的级

级	典型图像格式/像素	典型帧频/Hz	最大数据传输速率/(b/s)
L1	176×144	15	64k
L1b	176×144	15	128k
L1.1	352×288 或 176×144	7.5 或 30×192k	
L1.2	352×288	15	384k
L1.3	352×288	30	768k
L2	352×288	30	2M
L2.1	352×480 或 352×576	30 或 25	4M
L2.2	SD	15	4M
L3.0	SD	30 或 50	10M

(续)

级	典型图像格式/像素	典型帧频/Hz	最大数据传输速率/(b/s)
L3.1	1280×720	30	14M
L3.2	1280×720	60	20M
L4.0	1920×1080	30	20M
L4.1	1920×1080	30	50M
L4.2	1920×1080	60	50M
L5	2048×1024	72 或 30	135M
L5.1	4096×2048	30	240M

48.2 基本类和扩展类

读者也许认为基本类能被每一个解码器解码。在 H.264 中并非如此。基本类是用于处理会受低质量传输影响的低数据传输速率应用。灵活宏块排序 (FMO)、任意宏块条顺序 (ASO) 和冗余宏块条 (RS) 这些特性都有助于提高鲁棒性。为实现低计算复杂度，其他功能，特别是 B 宏块条，被排除在基本类之外。基本类很少用于（如果有使用）专业视频。

读者也许认为扩展类有超出主类的特性。在 H.264 中并非如此。扩展类通过包含数据分割和 SI 与 SP 宏块条这两个额外的特性，扩展了基本类的鲁棒性特点。B 宏块条和隔行编码 (PicAFF, MBAFF) 这两个附加的功能提高了编码效率。扩展类很少用于专业视频。

48.3 高类

原始的 H.264 特性由在高类中可用的保真度范围扩展 (FRExt) 进行了增强。

Hi10P 和 Hi422P 可用 10 位采样深度；Hi444P (高 4:4:4 预测类) 可用 14 位采样深度。

Hi422P 和 Hi444P 提供了 4:2:2 色度亚采样，可以编码 $Y'C_B C_R$ 4:2:2 (不严格地说， $Y'UV$ 4:2:2)。Hi444P 提供了 4:4:4 “色度采样”，也就是根本没有亚采样。

48.4 层次结构

H.264 码流中的句法元素有类似 MPEG-2 的层次结构。H.264 的位流层次结构——句法层次结构，如下：

- 序列；
- 图像；
- 宏块条；
- 宏块；
- 宏块分区；
- 子宏块分区；
- 块；
- 采样点。

视频编码层 (VCL) 包含宏块条及以下级的元素。网络抽象层 (NAL) 定义了承载编码数据的 NAL 单元。VCL 之上各层，也就是序列级和图像级的信息，用非 VCL 的

NAL 单元传输。两种类型的 NAL 单元（VCL 和非 VCL）可以在不同的流中传输，例如实现更好的网络鲁棒性，尽管这样的传输机制规范在 H.264 的范围之外。

附加增强信息（SEI）和视频可用性信息（VUI）是插入到编码位流的非 VCL NAL 单元中的“消息”。SEI 包含序列和图像参数集（SPS 和 PPS）。VUI 承载与 MPEG-2 的序列显示扩展内容相近的信息。

48.5 多参考图像

MPEG-2 有两个参考帧，一个在过去，一个在“将来”。“将来”帧可用于预测显示顺序在前的 B 图像。

在 H.264 中，允许多参考图像，数量在 4 个参考图像至 13 个参考图像之间，依赖于级别。如果正在编码的素材快速切换到一个反切镜头，编码器可以指示解码器保留第一个镜头结尾的图像，并用它来预测从反切镜头返回的图像。参考图像能够以任意顺序寻址。

多参考图像可能对于预测“未覆盖的背景”是有用的，依赖于编码器发现这种背景的能力。使用“将来”帧引起延迟，并且在一些应用中可能不现实。

48.6 宏块条

宏块条为解码器提供了并行的选项：帧内预测不会跨越宏块条边界。解码器状态实际上在宏块条的边界重置，所以宏块条限制了传输引起错误的空间范围。宏块条可进行冗余编码来进一步减少传输错误。

48.7 空间帧内预测

在 MPEG-2 中，一个宏块可以完全独立地编码为 I 宏块，或利用时间预测编码为 P 宏块。在 H.264 的开发中认识到，在当前宏块上方的解码得到的帧内宏块和同一宏块条中左边的宏块具有空间域预测价值。H.264 基于这些数据实现了帧内预测，在几种模式下直接复制上方或左边的图像数据。帧内预测接下来可通过常规方法用变换编码的量化后残差来细化。帧内预测只使用帧内编码宏块的信息。

还有一个帧内 PCM 模式，I 宏块像素数据不经过变换直接编码。该模式对于很高的数据传输速率可能有用。

48.8 灵活的运动补偿

在 MPEG-2 中，运动预测是以亮度像素的 16×16 块，即宏块为单位完成的。编码器试图找到参考图像中一个可形成很好的预测器的 16×16 区域，然后将该块的相对坐标作为运动向量编码到数据流中。

在 H.264 中，宏块可以划分成几种形状和尺寸，用于从参考图像的不同区域预测，甚至是从不同参考帧预测。整个宏块可以从一个 16×16 的源中预测；或者，宏块可以划分为两个 8×16 宏块分区，两个 16×8 分区，或 4 个 8×8 分区，各自独立预测。在高类中，如果宏块划分为 4 个 8×8 宏块分区，这些分区还能划分成两个 4×8 子宏块分区，两个 8×4 子分区，或四个 4×4 子分区，还是各自独立预测。宏块可以与多达 16 个运动向量相关联。

48.9 四分之一像素运动补偿插值

在 MPEG-2 中，运动向量相对于亮度采样点具有 $1/2$ 像素精度。在 H.264 中，运动

补偿插值能够以 1/4 像素精度执行，即运动向量能够以 1/4 像素精度为单位编码（有时称为 quarter-pel 或 Qpel）。插值操作使用简单的 6 抽头 FIR 滤波器，除了在最佳的空间位置产生预测信号外，还具有低通滤波的有益效果。

亮度的 1/4 像素精度就是 4:2:0 色度的 1/8 像素精度。

48.10 运动补偿预测的加权和偏置

在一幅图像渐变为另一幅图像，渐变为黑色，或在《六英尺之下》中渐变为白色时，MPEG-2 的性能很差。传输系数的直流分量编码还是不错，但是渐变为黑色时所有的交流项一起按比例减小了，这给量化带来很大压力。H.264 采用运动补偿预测的加权和偏置来提高在渐变和其他环境下的性能。

48.11 16 位整数变换

MPEG-2 像 JPEG 一样也使用 8×8 DCT，实际上总是用二进制定点运算来实现。理论上 DCT 矩阵包含无理数；编码器和解码器用定点二进制整数来近似表示这些数字，通常为 16 位。JPEG 和 MPEG-2 标准都没规定 DCT 的精度。编码器包含了解码过程的仿真，但是由于在不同的实现中舍入误差不同，编码器的 DCT 可能与解码器的 DCT 不匹配。当解码块用作预测时，解码器端形成的预测可能不会精确匹配编码器期望的预测。假定编码器比解码器有更多计算资源，而且很可能准确度更高，所以这个问题命名为解码器漂移——这么说也许对解码器不公平。

H.264 采用一种用简单的二进制分数定义的变换，其反变换也由简单的二进制分数构成，来消除解码器漂移。用 8 位残差和 16 位算术，不会产生舍入误差，也就不会产生漂移。

H.264 的变换有时用术语 HCT 表示，HCT 代表 H.264 余弦变换或高相关变换。

48.12 量化器

在 MPEG-2 中，变换系数量化电平是等间隔的。在 H.264 中，量化器有指数间隔的 52 级：每一级将步长增大到 1.122 倍，即六级就使步长翻倍（粗算起来，量化器步长增加 +1，数据传输速率减小大约 10%；步长增加 1 倍，则比特率减半。这种试探方法可用于编码器端码率控制）。

48.13 可变长编码

假设给出一个四符号 (A, B, C, D) 的序列要编码成位流。考虑表 48-3 给出的两个简单的编码方案。方案 F 对四个符号都用 2 位表示。方案 V 根据待编码的符号，使用 1 位、2 位或 3 位。两个方案都能可靠编码任何给出的输入序列，即两种编码是无损的。然而，如果输入包含许多个 A，方案 V 输出的位比方案 F 的少。方案 V 是可变长编码的基本概念的例证：能反映待编码符号概率的编码具有优越性。在这个例子中，方案 F 适用于 A、B、C、D 是等概率的输入。方案 V 适用于概率分别为 [1/2, 1/4, 1/8, 1/8] 的输入。

在 MPEG-2 中，设计有几十种 VLC 方案用于不同的句法元素。H.264 需要许多增加的语法元素，而开发者厌倦了建立特别的表格。一种

表 48-3 将符号 (A~D) 映射到比特流的两个假设的编码方案

符 号	方案 F	方案 V
A	00	0
B	01	10
C	10	110
D	11	111

系统化的方法，通用可变字长编码（UVLC）被采纳。UVLC 基于指数 Golomb 方案，表 48-4 给出了一个正数 1~11 或范围为 ± 5 的整数指数 Golomb 编码的例子。

对例子中的正数（POS）符号进行解码的过程如下：如果数据流位为 1，则编码的值是 1。否则，对先导零位计数，用计数值 N 表示。将接下来的 $N+1$ 位（包括开头的 1 位）看作二进制编码的正数，最高有效位在前。

表 48-4 中的 POS 例子帮助读者理解表的含义。H.264 的无符号整数（ue）编码比显示的数字少 1。表 48-4 的 INT 例子与 H.264 的有符号整数（se）编码相对应。

表 48-4 整数指数 Golomb 编码示例

正 数	整 数	编码的位流
1	0	1
2	+1	0 <u>10</u>
3	-1	0 <u>11</u>
4	+2	00 <u>100</u>
5	-2	00 <u>101</u>
6	+3	00 <u>110</u>
7	-3	00 <u>111</u>
8	+4	000 <u>1000</u>
9	-4	000 <u>1001</u>
10	+5	000 <u>1010</u>
11	-5	000 <u>1011</u>

当用于有符号整数（例子中的 INT 符号）时，解码如下：如果数据流的位是 1，则编码值是 0。否则，对先导零位计数，用计数值 N 表示。将接下来的 $N+1$ 位（包括开头的 1 位）看作编码数字的绝对值，以二进制表示，最高有效位在前。拖尾的第 $(N+1)$ 位为符号位。

表 48-4 的 INT 例子对像是那些在运动向量位移中遇到的有符号整数进行编码。通过简单地给需要的值或符号分配适当的数字，编码容易适应非数值符号。表 48-5 显示了编码怎样扩展到任意大的数字（或一组任意大小的符号）的。

表 48-5 推广的指数 Golomb 编码

整 数	编码的位流	整 数	编码的位流
0	1	$\pm 8, \dots, \pm 15$	0000 1 <u>xxxxs</u>
± 1	0 <u>1s</u>	$\pm 16, \dots, \pm 31$	00000 1 <u>xxxxxs</u>
$\pm 2, \dots, \pm 3$	00 <u>1xs</u>	$\pm 32, \dots, \pm 63$	000000 1 <u>xxxxxxs</u>
$\pm 4, \dots, \pm 7$	000 <u>1xxs</u>	$\pm 64, \dots, \pm 127$	0000000 1 <u>xxxxxxxxs</u>

注：指数 Golomb 编码可推广到用 1 位、2 位、3 位、4 位，直至无穷多位来表示有符号整数。该方案偏向于小数字最可能出现的输入：如果输入在 ± 127 内有同等可能性，则固定长度的 8 位二进制补码编码会更有效。

在 H.264 中，UVLC 在变换系数之上的句法层中使用，用于诸如预测模式和运动向量的数据。UVLC 方案不用于变换系数：CAVLC 或 CABAC 用于变换系数。

48.14 上下文自适应性

MPEG-2 的设计者们用他们的判断和经验以及许多实验结果来建立 MPEG-2 的 VLC

表。然而，这些表是静态的。

特定源素材的 VLC 项的用法可看作是一个统计分布，即 VLC 表的用法依赖于历史，依赖于上下文。

上下文自适应性指编码器动态跟踪表项的使用，估计它们使用的概率，并改变 VLC 的映射，以使编码的位流对很可能遇到的符号有紧凑的表示。

上下文自适应性造成编码器和解码器复杂度增加。在 H.264 中，上下文自适应性的基本形式是上下文自适应可变长编码（CAVLC），在宏块条级及以下级使用。

48.15 CABAC

当编码符号的概率可以表示为二进制分数时，上文概述的 VLC（及相关的 CAVLC）是最佳的，比如在表 48-3 所示方案 V 的例子中，[A, B, C, D] 的概率为 [1/2, 1/4, 1/8, 1/8]。一般而言，待编码符号的概率不会都与二进制分数相近。例如，也许要编码概率为 [1/3, 1/3, 1/3] 的三个符号。

一种叫做算术编码的技术可用于对每个符号占用相当于 1 位的分数的分布进行有效编码。采集到有可能很大的一组符号，然后编码为这组符号折合成的一个数字，而这个数字的范围划分为与单个符号的概率对应的子范围。

类似 VLC，算术编码可以是上下文自适应性的，从而有 CABAC；上下文自适应二进制算术编码。如果听起来复杂，CABAC 确实复杂。CABAC 可带来 10% 左右的比特率改善；然而，它增加了编码器和解码器的复杂度，且消耗处理器和存储器资源。CABAC 在 H.264 的主类和高类中可用。

CABAC 是 H.264 主类和高类的一部分；然而，H.264 所有权之外的企业可能会决定是否使用 CABAC。例如，CABAC 在 AVC-Intra 50 中允许使用，但在 AVC-Intra 100 中禁用。

48.16 去块滤波器

在 MPEG-2 中，反变换倾向于在两个 8×8 块邻接处产生间断点，即块效应，这是个问题。许多 MPEG-2 解码器加入后处理来减轻块效应的影响，但在产生块效应之后的处理（环路之外）对编码器不可见。

H.264 标准化了一种自适应的环路内去块滤波器。该滤器适应图像内容，如边缘（通常会造成最严重的伪影）。去块操作被标准化，并且在编码器的预测环路中进行。

48.17 缓冲区控制

在 MPEG-2 中，视频缓冲区校验（VBV）值从编码器传输到解码器，以确保解码器的缓冲区跟踪编码器的缓冲区，不出现下溢或上溢。

在 H.264 中，缓冲区控制更加复杂。在编码器和解码器端，有两个假想的缓冲区：编码图像缓冲区（CPB），表示编码的位流中的图像，以及解码图像缓冲区（DPB），表示解码后的图像，这两个缓冲区的内存占用都将被跟踪。

48.18 可分级视频编码

在 H.264 的附录 G 中定义的可分级视频编码（SVC）允许信息以一种分层级方式的结构传送，从而允许提取比特率低于完整序列的部分位流，允许解码有多个图像结构的图像（用空间可分级性编码的序列）、有多种帧频的图像（用时间可分级性编码的序列）和/

或有多级画质的图像（用信噪比/画质可分级性编码的序列）。

在单一的位流中，拥有有限的计算资源的解码器可以提取基本位流来解码出一个低级表示（在这种情况下，没有体现数据传输速率的优势）。

不同的层可以被分到不同的位流中。所有的解码器访问基本流，解码能力更强的解码器可以访问增强流。不过，对于某些应用，如 HTTP 直播流来说，以不同的速率编码同一个节目，每种速率在一个自我包含的流中，可能更有效，这样解码器可以简单地访问具有合适速率的单一的码流。

48.19 多视点视频编码

H.264 的附录 H 介绍的多视点视频编码（MVC）对用于有效编码两幅（或可能多于两幅）有高度空间相关性的图像的特性进行了标准化。常见应用是编码立体图像对的左、右图像。

MVC 增加了两个新的类，多视点高类（Multiview High Profile, MHP）和立体高类（Stereo High Profile, SHP）。用其中任何一个类编码的视频都与 H.264 高类后向兼容；解码器只看到基础视点（对于 SHP 的情况，基础视点一般是左眼）。

48.20 AVC-Intra

AVC-Intra 是松下对于使用 Hi10Intra 类的演播室质量 H.264 的表示法。不同帧频的 720p、1080i 或 1080p 格式的视频用 10 位 $Y' C_B C_R$ 4 : 2 : 2 分量表示。在 AVC-Intra 50 中，720p、1080i 和 1080p 视频经亚采样并将速率压缩到 50Mb/s。在 AVC-Intra 100 中，在原生像素数量下的 720p、1080i 和 1080p 视频将速率压缩到 100Mb/s，见表 48-6 所示的总结。

表 48-6 AVC-Intra 的类/级组合

类@级	高类 10 帧内 (Hi10Intra)
L3.2	AVC-Intra 50: CABAC, 4 : 2 : 2, 1280 × 720p 下采样到 960 × 720 像素
L4.0	AVC-Intra 50: CABAC, 4 : 2 : 2, 1920 × 1080p 下采样到 1440 × 1080 像素
L4.1	AVC-Intra 100: CAVLC, 4 : 2 : 2, 原生的 1280 × 720p and 1920 × 1080 像素

延伸阅读

- Richardson, Iain E. G. (2010), *The H.264 Advanced Video Compression Standard* (Chichester, U. K. : Wiley).
- Sullivan, Gary J., Pankaj N. Topiwala, and Ajay Luthra (2004), "The H.264/AVC advanced video coding standard: Overview and introduction to the fidelity range extensions," in *Proc. SPIE* **5558**: 454-474.
- Sullivan, Gary J. and Thomas Wiegand (2005), "Video compression-From concepts to the H.264/AVC standard," in *Proc. IEEE* **93** (1): 18-31.

第 49 章

VP8 压缩

2010 年，谷歌收购了一家名为 On2 的公司，On2 在 10 年或更长时间里开发出了一系列用于视频分发的基于软件的专有编解码器。谷歌将 VP8 编解码器开源，并将它作为一个用于将视频通过网络（基于 IP）分发到消费者的称为 WebM 提案的基础。WebM 由 VP8 编解码器编码的视频和 Vorbis 编解码器编码的音频组成，音视频都打包在 Matroska 文件封装器中。

VP8 遥远的前身 VP3 被 On2 公司开源。VP3 随后发展为 Theora。On2 将 VP6 和 VP7 授权给 Adobe 作为 Flash 8 视频的基础；随后，H.264 并入 Flash 9。On2 公司将 VP7 授权给 Skype。

VP8 编解码器大体上基于早先的章节中讨论的 MPEG-2 和 H.264 的原理，虽然谷歌打算 VP8 不受 MPEG-2 和 H.264 的知识产权（IPR，这里指专利权）的阻碍。VP8 的元素上的专利已发布给 On2；谷歌允许免许可费使用这些专利。谷歌对 VP8 的许可要求用户不诉讼任何针对 VP8 的知识产权（“互不主张”）。谷歌 VP8 的实现不侵犯不由谷歌控制的专利，也许甚至是 MPEG-2 或 H.264 的专利，这一点并没有担保或保障。

VP8 的技术和商业问题就是该描述性标准并不全面详细：VP8 的权威性规范实际就是其参考代码。在一些地方，含糊的代码引起 VP8 “标准” 应当根据其明显意图，还是根据编码执行的内容来定义这样的问题。由于缺少书面标准，实现者被迫把参考代码当作权威，即使性能或互操作性受到影响。

算法特性

如前面提到的，VP8 编解码器大体上以 MPEG-2 和 H.264 的原理为基础。为了理解下面的内容，读者应该熟悉第 16 章，以及前文介绍的 JPEG /M-JPEG、DV、MPEG-2 和 H.264。

VP8 只编码逐行、8 位、4:2:0 Y'C_BC_R 视频。没有关于隔行扫描的规定。

谷歌的文件用 YUV 指代 Y'C_BC_R。

VP8 包含称为关键帧的帧（相当于 MPEG-2 的 I 帧）和交互帧（像 MPEG-2 的 P 帧）。VP8 没有 B 帧，所有解码帧均可用于预测。VP8 解码器有三种参考帧：金质帧、先前帧和备用参考帧（altref 帧）。

每幅图像伴随着一个 1 位的 show_frame 标志，表示是否显示该帧。此标志可以使被解码的帧置入参考帧中但不显示。在不寻常的环境下，使用这种机制可以模仿 B 帧。

位流被划分成段。每一个段内有一个 4 字节的帧头和 1 到 9 个标记为 I, II, III 等的分区。分区是代表视频各方面的字节序列（类似 H.264 中 VCL NAL 单元和非 VCL NAL 单元分开）。分区 I 按光栅顺序传输每个宏块的预测模式和运动向量。I 之上的分区传输量化的变换系数（在 VP8 中，有时称为纹理）。宏块行可以映射到单一的分区，或能并行处理的 2、4、8 个分区（将要介绍的熵上下文在分区之间共享；二进制算术解码可以在一定程度上并行化，但编码不行）。

VP8 将 16×16 的宏块细分为 4×4 像素的子块。在每个 $Y'C_B C_R 4:2:0$ 宏块中有 24 个子块。不同于 H.264，VP8 没有 8×8 亮度块。色度预测是在 8×8 的色度块上进行。

VP8 有两种亮度帧内预测模式， $i16 \times 16$ 和 $i4 \times 4$ （没有 8×8 帧内亮度预测），参考同一帧中先前解码的像素。使用帧内预测就妨碍了并行性。

位流识别可获得每个块的帧内预测的四种方法之一。

- V_PRED 垂直预测 预测值从上面一行向下复制到块中。
- H_PRED 水平预测 预测值从左面一列复制到整个块中。
- DC_PRED 直流预测 预测值都被设置为上面一行和左面一列的平均值；这称为“直流”色度预测。
- TM_PRED 外差预测 预测值是用左上角的（固定）二阶插值对上面一行和左面一列外插得到的（此模式大致相当于 H.264 的平面预测）。

VP8 的核心变换是一个用 16 位整数系数近似的 4×4 DCT。解码器使用精确的 16 位算术，不存在解码器漂移。

对于 16×16 亮度预测模式，亮度处理涉及一个二级（Y2）变换：16 个亮度子块经 DCT 后，16 个直流系数集合起来进行一次（第 25 次） 4×4 变换。二级变换不是 DCT，而是沃尔什-哈达玛变换（Walsh-Hadamard transform, WHT）。

有六个量化器，每个都有各自的量化级。使用哪个量化器是依据“平面”（一阶亮度、二阶 [Y2] 亮度，或色度）以及系数是直流还是交流来确定的。

量化级是一个 7 位数字，索引量化表中的一个条目。量化潜在具有区域自适应性：编码器把每个宏块关联到四个类之一，每一类有一套不同的量化参数集。

VP8 实现了一个复杂的算术编码方案，比 CABAC 简单，但有类似的性能和较轻的处理负荷。编码器建立了对各种句法元素和参数值的概率估计。保持了默认的基准参数集；当出现关键帧时，概率分布被重置为基准。每一帧被处理时概率得到更新；编码器指示在解码完成时更新的参数集是成为新的基准（“持续的”）还是被丢弃（“一次性的”）。

VP8 有一个自适应环路内去块滤波器，其质量和复杂度与 H.264 的去块滤波器大体上可以比较。

延伸阅读

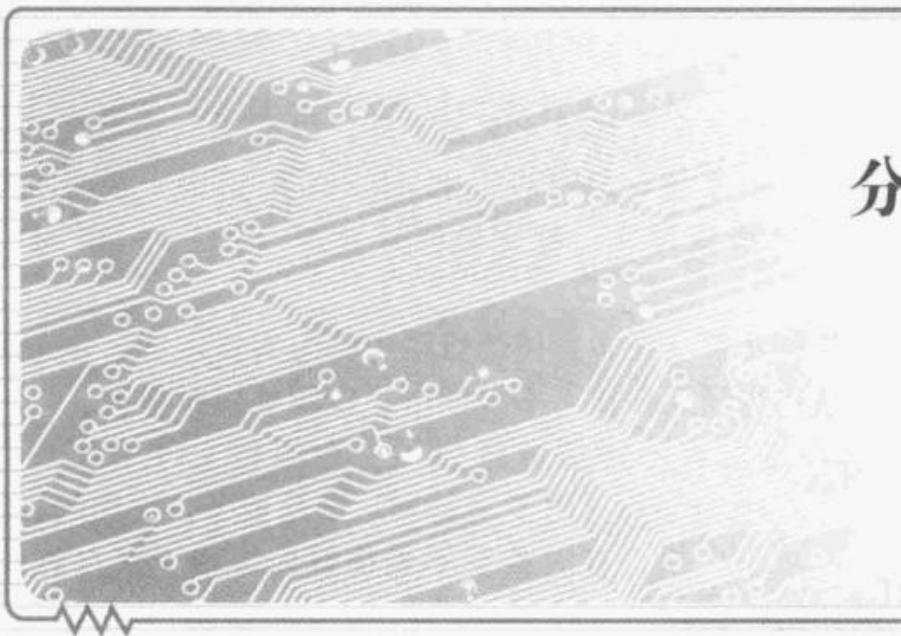
Bankoski, Jim, Paul Wilkins, and Yaowu Xu (2011), “Technical overview of VP8, an open source video codec for the web,” in *Multimedia and Expo (ICME), 2011 IEEE International Conf.* : 1-6.

Bankoski, Jim, Paul Wilkins, and Yaowu Xu (2011), *VP8 Data Format and Decoding Guide*, IETF Informational RFC. This information is available in a more readable form as Google On2 (2011), *VP8 Data Format and Decoding Guide* (revised 2011-02-04).

Feller, Christian, Juergen Wuenschmann, Thorsten Roll, and Albrecht Rothermel (2011), “The VP8 video codec-overview and comparison to H.264/AVC,” in *Consumer Electronics-Berlin (ICCE-Berlin), IEEE International Conf.* : 57-61.

第六部分

分发标准



第 50 章

MPEG-2 存储与传输

多媒体包含视频和音频，也可能伴有其他元素如字幕，以适于同步呈现给观众的方式编码。正在使用的有许多视频压缩系统；MPEG-2 和 H.264 广泛用于消费类应用。正在使用的有许多音频压缩系统；在消费领域，杜比数码环绕声（AC-3）和 MPEG-1 层Ⅲ（MP3）被广泛使用。

多媒体广播或分发要求各种元素（在多媒体的行话中称为实质内容）被复用进一个单一的文件或码流，其中的视频和音频元素接下来被同步，以能够同时呈现。

在多媒体计算中，复用是通过将不同成分排列到一个容器文件中实现的。微软的 AVI、苹果的 QuickTime，以及 Matroska（用于 WebM）都是例子。这样的容器格式相当适用于计算机，但通常不太适用于广播，有时甚至不能很好地适用于来自例如 DVD、蓝光光盘等介质的专用的、高性能回放。

一些在 PC 上使用的多媒体格式使用多个文件，例如一个文件用于视频，另一个文件用于音频。这类方案实际上将复用操作交给播放器软件。这类方案容易使某一类实质内容播放失败或使实质内容不同步。

MPEG-1 标准的系统部分从 1992 年起建立了一种复用结构。该方案延续到 MPEG-2 中，而且 MPEG-2 方案目前广泛用于计算机领域、广播及消费类视频应用（包括使用硬盘驱动器或闪存介质的消费类录像机）。MPEG-2 第 1 部分“系统”定义了两种复用机制，节目流（PS）和传送流（TS）。两种都视为 MPEG“容器”，其结构是本章其余部分的主题。

前面的 16.15 节简要地讨论了 ISO 基本媒体文件格式。该格式用作 MPEG-4 第 2 部分/ASP 视频容器格式。该格式普遍认为不适用于专业视频。

50.1 基本流

音频或视频编码器产生的字节流称为基本流（ES）。前一章概述了编码到视频基本码流中的信息（音频编码超出了本书范围）。

基本流可以包含私有流。

50.2 打包的基本流

基本流打包成 188 字节的包，第一个字节是 MPEG 的同步字节为 $(47)_{16}$ 。有的系统构建 204 字节的包，要求信道编码器覆盖每个包的最后 16 字节；在这种情况下，同步字节为 $(B8)_{16}$ 。

50.3 MPEG-2 节目流

MPEG-2 节目流（PS）是一个相对简单的复用单一节目的视频和音频的机制，用于在像是计算机磁盘或数字光学介质这些相对无差错的介质上存储或传输。PS 数据包长度可变；典型的是 1KB 或 2KB 的包，但一个包也可以长达 64KB。MPEG-2 节目流在这些

应用中使用：

- DVD 介质使用 MPEG-2 节目流编码的一个严格的子集；相关的文件扩展名是 .vob。
- MOD 消费类视频格式根据 DVD 的习惯基本上是一个 MPEG-2 MP@ML 标清节目流。在计算机上，这类文件通常有扩展名 .mpg 或 .mpeg。

50.4 MPEG-2 传送流

MPEG-2 传送流 (TS) 是 MPEG-2 系列标准的一部分，规定了一个相对复杂的将一个或多个节目的视频和音频复用到一个数据流中的机制，通常用短的数据包，适合于通过易出错而需要有相对强大的前向纠错 (FEC) 的介质传输。传送流适用于播放器连接到正在进行的传输（如电视），与从开头读取文件相对。对于地面无线下载 (OTA) 或有线电视，TS 包应当有适当保护；不过 FEC 和信道编码的标准在 MPEG 标准之外，通常在数字电视标准领域内（例如，北美的 ATSC 标准和欧洲的 DVB 标准）。

传送流数据包 (TSP) 由 188 字节构成，一个 4 字节头部（其第一个字节的值为 $(47)_h$ ），包括一个 13 位的包标识符 (PID)，以及 184 字节的载荷。数据包大小的设计考虑到了异步传输模式 (ATM)：一个 TS 包可以放到四个 ATM 信元（各 48 字节）中。由于缺乏节目流的外部接口，一个单节目传送流 (SPTS) 可用于携带一路节目。一些应用使用多节目传送流 (MPTS)。

ATM：异步传输模式，一种高性能联网协议。

PID 为 0 的 TS 包含有节目关联表 (PAT)，每秒重复若干次。PAT 列出一个或多个后续含有节目映射表 (PMT) 的包的 PID。一个 PMT 列出与单路节目相关联的视频和音频基本流的 PID。

ATSC 数字电视传送流中包含一组实现节目和系统信息协议 (PSIP) 的包。PSIP 识别频道和节目，并传输一天中的时间和电台呼号信息。PSIP 使接收器能够提供电子节目指南 (EPG)。

见 ATSC 标准 A/65。

在计算机上，188 字节的 TS 包一般附加一个 4 字节的时间码（形成 192 字节的数据包）；由一系列这样的包组成的文件一般具有扩展名 .m2t、.m2ts，或者只是 .ts。

MPEG-2 传送流在这些应用中使用：

- TOD 消费类视频格式（本质上是 MPEG-2 MP@HL 高清传送流）。
- 蓝光的 BDAV 容器。
- H.264 压缩视频。
- AVCHD 压缩视频（在计算机领域，文件扩展名常为 .mts）。

50.5 系统时钟

MPEG 的同步通过一个系统参考时钟 (SCR) 来实现。25Hz 和 29.97Hz（确切地说是 30/1.001）的最小公倍数为 30kHz；在 MPEG-2 中，90kHz 被选为节目参考时钟 (PCR) 的基准。节目时钟值用 33 位表示，足以提供超过 24h 的唯一的 PCR 值。

MPEG 系统定时基于一个 27MHz 的参考时钟，表示为给 PCR 增加了一个 9 位字段，取值为 0 到 299。表 50-1 列举了不同帧频下每帧的 PCR 计数值。

表 50-1 MPEG-2 每帧的 PCR 计数

帧频/Hz	每帧的 PCR 计数
30	3000
29.97	3003
25	3600
24	3750

27MHz 除以 90kHz 等于 300。

每个节目流有单一的参考时钟。MPTS 的不同节目可以有不同的节目时钟，因此针对传输多个独立的 PCR 的传送流做了规定。

延伸阅读

Chen, Xuemin (2002), *Digital Video Transport System* (Springer) .

Whitaker, Jerry C. (2003), “DTV Service Multiplex and Transport Systems,” Chapter 13.2 in *Standard Handbook of Video and Television Engineering*, Fourth Edition (McGraw-Hill) .

Whitaker, Jerry C. (2003), “DTV Program and System Information Protocol,” Chapter 13.4 in *Standard Handbook of Video and Television Engineering*, Fourth Edition (McGraw-Hill) .

第 51 章

数字电视广播

本章简要总结数字电视广播。大部分标准化的数字广播系统都基于 MPEG-2 压缩而形成，在“MPEG-2 视频压缩”中介绍。一些有线电视和卫星系统使用 H.264 标准。

高清电视（HDTV）传输系统按设想能在一个 6MHz 的模拟频道上传输垂直分辨率和水平分辨率都约为标清电视（SDTV）2 倍的图像，即约 200 万像素。MPEG-2 可将每秒 30 帧的 200 万像素的图像压缩到大约 20Mb/s。适于地面射频传输的现代数字调制方案在信道带宽上有约为 3.5b/Hz 的净荷。结合这些数字可以看出，一路 HDTV 数字信号可以在原来被一个模拟 NTSC 6MHz 频道占用的频谱中发送。

自从 1941 年黑白电视出现，525 行、60 场/s 的隔行传输方案的基本射频参数基本上就没变。由于对其他频道的潜在干扰，调制方案要求在许多位置上的潜在频道保持不被使用。未使用的频道称为禁忌频道。数字电视传输利用了调制系统半个世纪的技术改进。选中的调制系统允许使用非常低的功率。这种低功率有两个主要后果：使从数字发射机到 NTSC 或 PAL 的干扰最小化，并且允许数字电视发送使用原先的禁忌频道。数字电视服务就这样叠加在模拟服务之上（在早期的 HDTV 部署中，节目素材在传统的模拟发射机上同时播出）。

51.1 日本

日本在 20 世纪 90 年代早期就部署了基于 1035i30.00 扫描和 MUSE 压缩的 HDTV 广播，该系统称为 Hi-Vision。MUSE 是一种为直播卫星（DBS）做了优化的模拟/数字混合系统；它记载于 NHK 实验室的书中。MUSE 早于 MPEG 标准；现在普遍认为，日本过早地采纳了模拟 HDTV 传输标准。

见 NHK 科学和技术研究实验室（1993 年）的《High Definition Television：Hi-Vision Technology》（New York：Van Nostrand Reinhold）。

在 2003 年，MUSE 被 ISDB-T 取代，ISDB-T 是一个基于 MPEG-2 视频编码和 OFDM 传输的地面广播系统。

51.2 美国

美国的高清电视开发者计划过让广播商用 HDTV 逐步取代模拟 SDTV 的传输。在 HDTV 发展的中途，情况变得明朗——同样的允许一个 HDTV 频道以大约 20Mb/s 编码和传输的压缩与传输技术，也一样适用于允许 5 个 SDTV 频道编码、复用并分别以 4Mb/s 传输！就这样，一开始是高清电视演变为包括 SDTV 和 HDTV 的数字电视（DTV）。压缩遵照 MPEG-2 MP@ML 用于标清，MP@HL 用于高清，有一些 ATSC A/53 规定的限制。

美国的数字电视标准由先进电视系统委员会（ATSC）制定。这些标准被联邦通信委员会（FCC）采纳，有一处显著的更改：FCC 驳回了记载于 ATSC A/53 标准表 3（见本书表 15-1）中的一组 18 种格式。迫于计算机产业的压力，FCC 删除该表，但没改动 ATSC 标准其余部分。表 3 不见诸于 FCC 标准实际上并不重要：在实践中，消费类接收

机不得不解码表 3 中的格式，不能依赖它们去解码任何别的格式。

见 ATSC A/53《数字电视标准》。

据说 FCC 删除表 3 是把光栅标准的选择留给市场：原则上，任何遵从 MPEG-2 MP@HL 的格式均可以使用。实际上，没有美国广播商选择，也没有任何消费类设备能确保实现 ATSC 表 3 以外的任何格式。实际上，数字电视解码器符合 MPEG 标准，并符合 ATSC 标准施加的限制。

在美国，数字电视音频用杜比数字音频编码（也称为 AC-3）标准化，一种在 MPEG-2 标准中没有规定的编码方案。杜比数字支持“5.1”声道：左和右（立体声）声道、前中央声道，左和右环绕声道，以及一个用于连接到“低音炮”的低频效果（LFE）声道（即符号中的“.1”）。ATSC 音频最大占用 512kb/s。

EIA-708-B（《数字电视（DTV）隐匿字幕》）标准化了传输数字电视隐匿字幕数据（DTVCC）的方法。

一个或多个 MPEG-2 节目流、相关的音频流、辅助数据和其他数据复用成数据传输速率约 19.28Mb/s 的传送流。传送流增加了 Reed-Solomon 前向纠错：每个 188 字节的传送包加上 16 个 FEC 字节，产生 204 字节的数据包传送给调制器。

51.3 ATSC 调制

ATSC 标准化了 8 电平数码残留边带（8-VSB）调制，每秒发送约 1076.2 万个 3 位符号。为使接收器能克服传输中引入的错误，两个前向纠错（FEC）方案串联使用：外编码是 Reed-Solomon（RS）编码，内编码是一个简单的 3/2 格型码。在 RS 和格型编码阶段之间，数据进行交织。在交织和编码后，加入含有段和场同步的同步信息。一个低级导频载波插到较低的频带边沿之上 310kHz 处，以辅助载波恢复。模拟技术用于上变频到 UHF 广播频道。

在接收机端，模拟技术将 UHF 广播下变频到中频（IF），通常是 44MHz。然后实现数字解调。通常，一个模拟的频率和相位锁定环路（FPLL）基于导频载波恢复载波频率。接下来，正交解调器恢复 I 和 Q 分量。I 分量以 10.76MHz 进行模/数转换恢复位流；对 Q 分量进行处理，以实现相位控制。位流然后经过格型解码、去交织、RS 解码和 MPEG-2 解复用。

设计一个不受多径失真、来自 NTSC 发射机的同频道干扰等传输损伤影响的解调器是一项挑战。干扰抑制滤波器，一种梳状滤波器的变体，内置于解调器中；滤波器削弱了可能有干扰的 NTSC 信号的视频、色度和音频载波。内置于解调器的自适应均衡器减轻了多径失真；信号的场同步分量用作参考信号。自适应均衡器通常用 FIR 滤波器实现，其系数作为估计的信道参数的函数被动态更新。

有线电视的信道特征与地面广播相比大不相同。有线 DTV 通常不使用 8-VSB 调制，而使用正交幅度调制（QAM），有 64 或 256 个电平（64-QAM 或 256-QAM）。

ATSC 定义了一种不用格型编码的 16-VSB 有线模式，但这种模式没有被部署。

对于直播卫星（DBS），正交相移键控（QPSK）被广泛使用。

美国的消费类接收机必须接受 ATSC 表 3（本书表 15-1）中不同的帧频和光栅标准。尽管多重扫描显示器在计算机领域很普及，当显示器不得不支持多种速率时价格和性能都会受到影响。多数消费类 HDTV 接收机的显示器工作在有限的扫描标准之上；ATSC 表 3 的宽范围是通过数字重采样来支持的。在 HDTV 早期部署中，许多接收机使用 1366×768 显示器，将 720p 上转换或 1080i 下转换到此格式。现在，大多数消费类 HDTV 接收机是

原生的 1080i，将其他格式转换为 1080i。

51.4 欧洲

欧洲在 20 世纪 80 年代和 90 年代花费了巨大的努力开发一种 HDTV 广播系统，其使用 1250/50 扫描以及一个建立在起初为 576i 设计的 MAC 传输技术之上的传输系统。MAC 在市场中失败了。HD-MAC 也失败了；部分是 MAC 商业失败的后果，部分是由于 HD-MAC 的技术弱点，部分是因为 HD-MAC 没有针对全球范围的市场。

欧洲的 HDTV 广播来得晚了，但基于 MPEG-2 MP@ML，有 720×576 像素图像结构的标清数字广播得到部署。数字视频广播（DVB）组织创建了一套全面的、用于有线（DVB-C）、卫星（DVB-S）和地面（DVB-T）广播的标准。DVB 音频遵从 MPEG-2 音频。这些标准由 ETSI (www.etsi.org) 颁布。

DVB-T 选中的射频调制系统是编码正交频分复用（COFDM）。COFDM 使用大量的子载波以将信号的信息内容均匀扩展到整个信道上。COFDM 的子载波单独调制，通常使用 QPSK 或 QAM。与 8-VSB 相比，COFDM 展示了大大提高的抗多径失真的能力。COFDM 还支持单频网络（SFN）传输，同样的位流从不同地点的多个发射机发送，但运行在相同的频率上。

延伸阅读

- Adams, Michael, (2000), *Open cable Architecture* (Indianapolis, Indiana, U. S. A. : Cisco Press) .
- ATSC A/54A, (2006), *Recommended Practice: Guide to the Use of the ATSC Digital Television Standard, including Corrigendum No. 1* (Dec.)
- Ciciora, Walter, James Farmer, David Large, and Michael Adams (2004), *Modern Cable Television Technology*, Second Edition (San Francisco: Morgan Kaufmann) .
- Collins, Gerald W. (2001), *Fundamentals of Digital Television Transmission* (New York: Wiley) .
- Dambacher, Paul (1998), *Digital Terrestrial Television Broadcasting* (Berlin: Springer) .
- Fischer, Walter (2008), *Digital Video and Audio Broadcasting Technology: A Practical Engineering Guide*, Second Edition (Berlin: Springer) .
- Reimers, Ulrich, ed. (2005), *DVB: The Family of International Standards for Digital Video Broadcasting* (Berlin: Springer) .
- Weiss, S. Merrill (1997), *Issues in Advanced Television Technology* (Boston: Focal Press) .
- Whitaker, Jerry C. (2003), “The ATSC DTV system,” Chapter 13.1 in *Standard Handbook of Video and Television Engineering*, Fourth Edition (McGraw-Hill) .

附录 A

使用 YUV 和照度是有害的

这是一个对精确使用术语的请求。标记 YUV 和术语照度在数字视频、计算机图形、数字图像处理中广泛传播。实际上，数字视频几乎从来不使用 $Y'UV$ 色差分量，也从不直接表示色彩学的照度。常见术语几乎总是错的。这份笔记解释了为什么。作者敦促视频工程师和计算机图像专家使用正确的术语，几乎总是 $Y'C_B C_R$ 和亮度。

A. 1 水泥与混凝土

作者用类比来展示为什么使用正确的术语非常重要。下次当你排队等公交车时，问你旁边的人什么建筑材料用于建造人行道。很可能那个人会回答“水泥”。

正确的答案是混凝土。水泥是石灰和黏土，为细腻的灰色粉末形式。水泥是混凝土的一种成分，其他成分是砂子、砾石和水。

平时你不需要精确使用这些术语：如果你提到人行道是用“水泥”建成的，人们会懂你的意思。外行人不会被术语水泥弄迷糊。有趣的是，专家们也不会困惑。如果监工头对工头大喊“给我 50 磅[⊖]水泥！”工头立即能从上下文中明白监工是否实际想要混凝土。然而，如果你给当地的建材供应商打电话订购 50 磅水泥，你一定不会收到 50 磅混凝土！外行人不会遇到不严格术语的麻烦，专家们也没有多少麻烦。介于二者之间的人容易被困惑。更糟糕的是，他们容易使用模棱两可或错误的术语，自己却没有意识到！

A. 2 真正的 CIE 照度

色彩学原理规定真正的 CIE 相对照度，记为 Y ，是线性（三刺激色值）RGB 分量的加权和。假如 CIE 照度在视频系统中传输，该系统将会符合恒定照度原理。可是在视频中，我们实现了一种背离了这个原则的工程近似。它在 1953 年为 NTSC 进行了标准化，并仍然是所有现代视频系统（标清和高清）的标准，亮度，记为 Y' ，是非线性（经伽马校正的） $R'G'B'$ 分量的加权和：

$${}^{601}Y' = 0.299R' + 0.587G' + 0.114B' \quad (\text{A-1})$$

绝对照度，用符号 L 代表，单位为 $\text{cd} \cdot \text{m}^{-2}$ （俗称 nit），很少用于视频。我们一般使用一个相对照度的近似值，用符号 Y 表示。

通常使用的非线性传递函数大致相当于一个平方根。我们利用色彩学的理论系数，但我们在一个不同于色彩学规定的框图中使用它们：伽马校正是在形成加权和之前进行，而不是之后。操作顺序与你从色彩学中预计的顺序是颠倒的。

A. 3 对照度的误解

20 世纪 50 年代，视频工程师认识到视频数量 Y' 与 CIE 照度大不相同，所以需要将其与照度区别开。他们使用词组“照度的数量表示”或“照度信号”来描述。他们用符号 Y

[⊖] 1 磅 = 0.4535kg。——编辑注

表示，但是增加了撇号来标记非线性： Y' 。显然，修饰语“数量表示”很冗长，而且经过几十年，它被省略了。随着时间的推移，撇号也被省略了。

不幸的是，没有发明新词来补充照度，强化色彩学数量和视频数量的区别。如今的多数视频工程师不熟悉色彩学，多数人不明白这个区别。工程师现在常常不小心用照度这个词和符号 Y 指代非线性（伽马校正后的） $R'G'B'$ 分量的加权和。

马虎的命名法进入了看起来权威的视频参考文献中，如 Pritchard 发表于 1977 年的有影响力的 SMPTE 论文。

计算机图形学先驱 Alvy Ray Smith 在使视频原理适应计算机图像学的探索中遇到了照度这个词。显然 Smith 将术语照度的使用与他的色彩学知识关联起来，并得出了视频“照度”和色彩学的照度相同的结论，这种想法尽管是错误的，但可以理解。因此，视频 $Y'IQ$ 被引入计算机图形学，并称其 Y 分量与 CIE 照度相同。

参阅如下文献。

Pritchard, Dalton H. (1977), “U. S. color television fundamentals-A review,” in *SMPTE Journal*, **86** (11): 819-828 (Nov.) .

Smith, Alvy Ray (1978), “Color gamut transform pairs,” in *Computer Graphics* **12** (2): 12-19 (Aug., Proc. SIGGRAPH) .

Foley, James D., and Andries van Dam (1984), *Fundamentals of Interactive Computer Graphics* (Reading, Mass. : Addison-Wesley) .

Foley, James D., Andries van Dam, Steven Feiner, and John Hughes (1990), *Computer Graphics: Principles and Practice*, Second Edition (Reading, Mass. : Addison-Wesley) . 589 (Section 13.3.3) .

不正确的解释传播到权威的计算机图形学教材中。《Computer Graphics: Principles and Practice》第 2 版，在标题为“YIQ Color Model”一节，有如下句子：

YIQ 的 Y 分量不是黄色，而是照度，并定义为与 CIE Y 基色相同。

加粗体强调出自原文。“黄色”指 CMY；在前一节提到了印刷油墨。“CIE Y 基色”用“CIE Y 分量”表示会更准确。

与引用的段落相反，所谓的视频 Y 分量，加上撇号表示为 Y' 更恰当，与 CIE 照度不同。视频 Y' 甚至不能从 CIE Y 计算得到，除非还有其他两个颜色分量（通常为基于 $B'-Y'$ 和 $R'-Y'$ 的色差分量）。引用的段落大错特错。显然，几十万份不同版本的副本，以及这本书的改编版已被印刷，混淆已经蔓延。

错误传播到数字图像处理社群。Pratt 的教材声明：

NTSC 为传输三刺激值 YIQ 制定了颜色坐标系统。Y 三刺激值是颜色的照度。

见 Pratt, William K. (1991), *Digital Image Processing*, Second Edition (New York: Wiley), 64. The error is corrected in the third [2001] and fourth [2007] editions.

视频数量确定不是三刺激值，三刺激值按 CIE 的定义，正比于强度。

视频工程师这方面马虎的命名法误导了一代数字图像处理、计算机软件和计算机硬件工程师。

A.4 亮度的奉行

术语亮度 (luma) 在视频中使用了很长一段时间，并没有一个精确的解释。为了采纳术语亮度来表示非线性视频的数量，作者在视频工程师之间和计算机图形学专家之间发起活动。该术语对视频工程师来说没有阻碍，实际上它比照度更顺口。由于它是与照度不同

的单词，亮度这个词吸引来自其他领域的读者在关于亮度和照度的关系下结论之前进行全面研究。

在 Fred Kolb 的帮助下，作者的活动成功了：1993 年，SMPTE 采纳了工程指引 EG 28《Annotated Glossary of Essential Terms for Electronic Production》。EG 28 定义了术语亮度，并澄清术语照度的两个冲突的解释。尽管 SMPTE EG 不算是 SMPTE “标准”，总算这个术语已经得到官方认可。视频工程领域论文和书籍的作者不再有任何借口马虎地使用术语照度。假使术语亮度普及于 20 年前，当 A. R. Smith 正在写关于 YIQ 的内容，或 Foley 和 van Dam 正在准备《Computer Graphics: Principles and Practice》时，这一团糟本来是可以避免的。可是那时还没有 EG 28。

如今 SMPTE、ISO/IEC、ITU-R 和 ITU-T 标准坚持用不正确的词语照度，从不提及与色彩学的 CIE 标准的歧义甚至是冲突，这真是耻辱。

A.5 色差比例因子

为了表示颜色，亮度伴随着两个色差，或色度分量，一般是基于蓝色减去亮度和红色减去亮度，其中蓝色、红色、亮度都经过伽马校正： $B'-Y'$ 和 $R'-Y'$ 。不同应用对基本的 $B'-Y'$ 和 $R'-Y'$ 分量使用不同的比例因子。 $Y'P_B P_R$ 比例因子为模拟分量视频进行了优化。 $Y'C_B C_R$ 的比例因子为数字分量视频进行了优化，如 4 : 2 : 2 演播室视频、JPEG 和 MPEG。 $Y'UV$ 和 $Y'IQ$ 比例因子的正确使用限于形成复合 NTSC 和 PAL 视频。

此处提到 NTSC 和 PAL 时，指的是彩色编码而不是扫描。不是指 480i 与 576i，或 525/59.94 与 625/50！

BT. 601 定义的 $Y'C_B C_R$ 缩放对分量数字视频来说是合适的。 $Y'C_B C_R$ 色度几乎总是使用三个方案之一进行亚采样：4 : 2 : 2 或 4 : 2 : 0 或 4 : 1 : 1。

见 ITU-R Rec. BT. 601-5《标准 4 : 3 和宽屏幕 16 : 9 帧型比数字电视的演播室编码参数》。

$Y'UV$ 缩放只适合用于形成复合 NTSC 或 PAL 制视频信号的中间步骤。当各分量是分开的时， $Y'UV$ 缩放就不恰当。然而， $Y'UV$ 命名的使用相当马虎，有些时候，特别是计算时，它指任何 $B'-Y'$ 和 $R'-Y'$ 的缩放。

大约在 1991 年，能用 IP 协议跨以太网传输文件的数字硬盘录像机（DDR）推出。Abekas 给这些文件引入了扩展名 .yuv。不过典型使用的（后来和现在）比例因子实际上对应 $Y'C_B C_R$ 。.yuv 扩展名的使用增加了 YUV 命名的误导性。

亚采样是一种数字技术，只对数字分量视频执行，即在 $Y'C_B C_R$ 上执行。在除了一些很专门的应用（即，4f_{SC} NTSC 或 PAL 复合视频的数字编码）中，亚采样不适用于 $Y'UV$ 。如果你看到描述为 $Y'UV$ 4 : 2 : 2 的系统，你会左右为难。也许写说明的人并不熟悉分量视频原理，设备（或软件）实际实现的比例因子是正确的。不过你必须允许有设计或实现该系统的工程师使用了错误的比例因子的可能性。如果使用了错误的公式，那么色彩精度会受影响，而且还难以诊断。

色度分量的正确顺序是先 $B'-Y'$ 然后是 $R'-Y'$ ，或先 C_B 后 C_R 。蓝色与 U 相关，红色与 V 相关；U 和 V 按字母顺序排列。

适当的 $Y'C_B C_R$ 缩放常用于 M-JPEG 和 MPEG。然而，用于计算机应用的静止帧 JPEG/JFIF 的 $Y'C_B C_R$ 缩放，通常使用全部的亮度和色度幅度，没有任何上余量或下余量。色度幅度是亮度幅度的 256/255。该缩放几乎就是 $Y'P_B P_R$ 的缩放，但不幸的是描述为 $Y'C_B C_R$ ：现在连 $Y'C_B C_R$ 都有歧义！把正确的 $Y'C_B C_R$ 缩放集成到 JFIF 为时已晚；在计

算机领域压缩的静态帧和运动图像库势必要遭受转换过程。

$Y'IQ$ 编码在演播室的使用已过时至少 30 年。现在应该抛弃 $Y'IQ$ 支持 $Y'C_B C_R$ 。

A.6 结论：澄清

对视频 Y' 使用照度这个词无异于用词语水泥而不是混凝土来描述人行道的主要建筑材料。外行人不关心，专家可以忍受，但夹在两者中间的人——在这种情况下，指在计算机领域重新实现视频技术的程序员和工程师，易于从粗心使用的术语得出错误的结论，从而产生不准确的颜色。图像的准确交换受到损害，用户遭受损失。

作者敦促视频工程师和计算机图形学专家避免使用术语 YUV、 $Y'UV$ 、YIQ、 $Y'IQ$ 和照度，除非在高度专业的情况下这些术语在技术上是正确的。恰当的术语几乎总是 $Y'C_B C_R$ 和亮度。

附录 B

辐射测定与光度测定简介

辐射测定领域涉及光功率及其空间与角度分布。光度测定本质上是用视觉谱响应加权的辐射测定。这两个领域涉及几个微妙的概念，被一堆令人困惑的符号和单位所掩盖。作者尽力理清一些让人困惑的地方，并给出一些关于单位和命名法的建议。

表 B-1 总结了辐射测定量、符号、单位（在左面的列中），及相应的光度测定量、符号、单位（在右面的列中）。光度测定量的符号就是对应的辐射测定量的符号，加上了下标 v （表示视觉）。有些人将下标 e 加到辐射测定符号上。符号 L_v （或 L ）用于绝对照度（例如在光度测定中）。符号 Y 用于相对照度（例如在色彩学和视频中）。

见 ANSI/IESNA RP-16。

CIE N° 17.4 (E-1.1) (1987), *International Lighting Vocabulary*, 4th Edition (Vienna, Austria: Commission Internationale de L'Éclairage).

表 B-1 辐射测定和光度测定的量、符号、单位

相对于以下量 的通量微分	辐射测定量 (辐射的)		光度测定量 (发光的)	
	数量 (符号)	单 位	数量 (符号)	单 位
—	辐射通量, 功率 (ϕ , F , P)	W	光通量 (ϕ_v , F_v , P_v)	lm
面积	辐照度 (E), 辐射出度 (M)	$W \cdot m^{-2}$	光照度 (E_v), 光出度 (M_v)	$lm \cdot m^{-2} = lx$
立体角	辐射强度 (I)	$W \cdot sr^{-1}$	发光强度 (I_v)	$lm \cdot sr^{-1} = cd$
面积和立体角	辐射率 (L)	$W \cdot sr^{-1} \cdot m^{-2}$	辉度 (L_v)	$cd \cdot m^{-2} = nt$

辐射测定和光度测定涉及空间中的光。不需要表面！光照特性可以在一个真实的或假想的表面描述；然而，光照特性并不是表面的特性。吸收率 (α)、反射率 (ρ) 和透射率 (τ) 是表面的固有特性，而不是光的特性。

后文按通常的物理习惯，用斜体书写字母符号，用罗马字体书写单位。

B. 1 辐射测定

辐射测定从测量波长在大约 100nm 与 1mm 之间的能量（使用符号 Q ）开始。能量使用单位焦耳 [J]。光子的能量 (Q_p) 与其波长 (λ) 相关，这里以 m 为单位给出：

$$Q_p = h \frac{c}{\lambda} \quad (B-1)$$

式中： h 为普朗克常数， $h \approx 6.626 \ 075 \ 5 \cdot 10^{-34} J \cdot s$

c 为光速， $c \equiv 299 \ 792 \ 458 m \cdot s^{-1}$

流速，或正式说法是，辐射能量的时间导数，是功率 (P)，也称为辐射通量 (F ，或用 Φ 更好)，以瓦 (W) 为单位。

每单位面积的辐射通量，即通量密度，穿过半球的所有方向到达一个点，就是辐照度 (E)。辐照度的单位是瓦每平方米 ($W \cdot m^{-2}$ 或 W/m^2)。正午时阳光辐照度 (日照) 约为

$1\text{kW} \cdot \text{m}^{-2}$ 。

辐射出射度把发射光和反射光加起来。前者的术语发射度排除了反射光。一些热学工程师用术语辐射率表示辐射出射度。在计算机图形学中，辐射率指的是一种计算照度的专业技术，并不指任何特定量。

在所有方向上离开一个点的单位面积辐射通量就是辐射出射度 (M)。正式说法是，相对于面积的辐射通量的倒数。辐射出射度以瓦每平方米 ($\text{W} \cdot \text{m}^{-2}$) 为单位。来自非发光表面的辐射出射度就是简单地把其辐照度乘以其反射率。

在一个特定方向上的辐射通量，正式说法是，每单位立体角的辐射通量，就是辐射强度 (I)；它的 SI 单位是瓦每球面度 ($\text{W} \cdot \text{sr}^{-1}$ 或 W/sr)。强度必须指定在一个特定方向上。强度是点状光源的属性，它独立于与观察者或测量仪器的距离。强度不遵循平方反比定律！强度在物理学中有许多其他含义，最常见的是单位面积功率，但这个量明确的术语是辐照度。

在摄影学中，符号 I 常用于表示辐照度或光照度，而不是强度。当心，声音强度在概念上不同于光强度：声音强度的量度是每单位面积的功率，与光的辐照度类似。

已故的 James M. Palmer 指出术语强度没有充分理由地被广泛滥用，因为它是 SI 系统的七个基本单位之一！

当心，不是以单位瓦每球面度 ($\text{W} \cdot \text{sr}^{-1}$) 表示的所谓强度。有些热学工程领域的作者，以及有些关于辐射率的计算机图形学领域的作者用术语辐射强度来表示辐射率。现在就来讲辐射率。

在一个特定方向上的辐射通量密度就是辐射率（符号 L ）。正式地，辐射率是辐射通量对立体角和投影面积的微分；图 B-1 给出了该定义的几何学表示。辐射率用单位瓦特每球面度平方米 ($\text{W}/(\text{sr} \cdot \text{m}^2)$ 或用 $\text{W} \cdot \text{sr}^{-1} \cdot \text{m}^{-2}$ 更好) 表示。对于一个大型非点光源，辐射率与距离无关。

图 B-1 中， dA 表示单位面积； $d\omega$ 表示单位立体角。投影面积收缩成 $\cos \theta$ 。

奇怪的是，在一些领域中，尤其是热传输、天体物理学和天文学中，辐射率称为强度。另外，今天称为辐射率的曾称为辐射测定亮度或就是亮度。术语亮度仍在天文学中使用，但在图像和色彩学中已被废弃。

辐射率可看作辐射测定的基本量：所有其他辐射测定量都能从它计算得到。读者可能认为从辐射率开始介绍比较直观，那么考虑以下内容：

- 辐射强度是辐照度在一块面积的积分。
- 辐照度是穿过立体角的辐射率的积分，即在半球所有方向上的积分。
- 通量是穿过一块面积的辐照度的积分，或等效地，穿过立体角的辐射强度的积分。

所有这些辐射测定术语与宽频谱的波长有关。其中任何术语都可以限定在一个窄的频谱，需要在术语前加上“光谱的 (spectral)”，给字母符号添加下标 λ ，并在单位后加上每纳米 ($\cdot \text{nm}^{-1}$)。

B.2 光度测定

到目前为止，已经讨论了辐射测定的物理量。光度测定完全相似，除了每个量的光谱

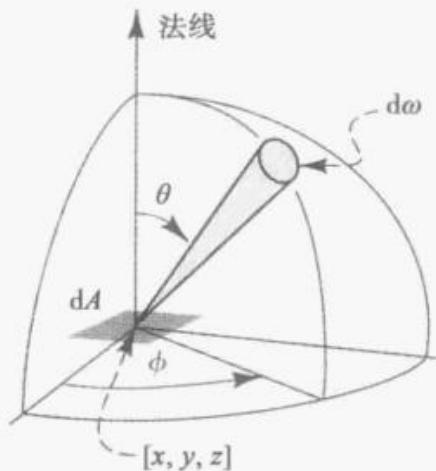


图 B-1 与辐射率定义相关的几何学表示

成分要用人眼视觉的光谱灵敏度加权，人眼视觉的光谱灵敏度标准化为 CIE 标准观察者的光照效率（见图 20-1）。

辐射测定和光度测定通过坎德拉的定义联系在一起：1 坎德拉 [cd] 是指具有 $1/683 \text{ W} \cdot \text{sr}^{-1}$ 辐射强度的单色 540THz 光源的发光强度。一旦这个定义建立起来，剩下的光度测定量和单位就平行于辐射测定量和单位。其联系如图 B-2 所示，该图顶部的四边形显示了辐射测定量：辐射通量、辐照度、辐射强度和辐射率。在底部的是光度测定量：光通量、光照度、发光强度和照度。两个系统通过坎德拉用辐射强度术语的定义联系起来。

1lm 由功率约 1.5mW 的 555.5nm 单色光产生。此波长，频率为 540THz，对应视觉峰值光照效率。

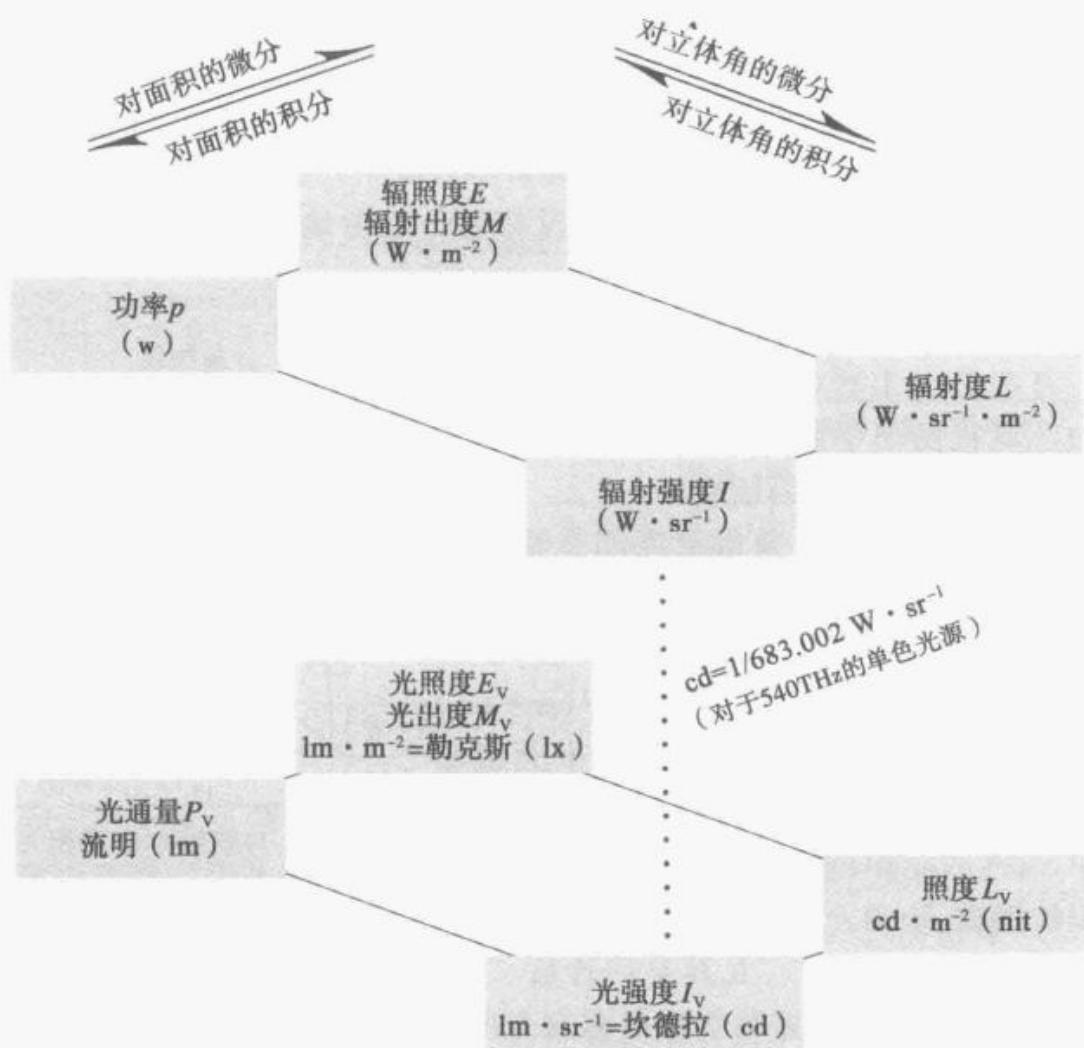


图 B-2 辐射测定量和光度测定量及其单位的联系

辐射通量在光度测定的类似量是光通量 (Φ_V)。用 James Palmer 的话说，光通量是当你买一个灯泡时你想要的。其亮度单位是流明 (lm)，是瓦在光度测定的类似量，其效率用流明每瓦 (lm/W) 测量。不管其光谱组成如何，一流明看起来亮度相同。

到达一点的每单位面积光通量，即光通量密度，就是光照度 (E_V)，SI 单位是勒克斯 [lx]。1lx 被定义为 $1\text{lm} \cdot \text{m}^{-2}$ 。光照度是辐照度在光度测定的类似量；这个量是用入射光计测量的。离开一点的每单位面积光通量就是光出射度 (M_V)。不论光是射入还是射出，然而，传统派常用单位 $\text{lm} \cdot \text{m}^{-2}$ 表示光出射度。不发光表面的光出射度是其光照度乘以其反射率。

光照度以前称为照明。不过，已经不再用照明表示这个量，因为这个词更普遍的意义为照明这个动作，或正被照亮的状态。

老的光出射度单位英尺烛光和米烛光有误导性，因为老的烛光是光强度单位，而不是

光通量单位。

在一个特别指定方向上的光通量就是发光强度 (I_v)，单位为 lm/sr 或 cd 。坎德拉是与老的烛光（口语化称为单位烛光）单位等效的现代单位。强度独立于距离：一支蜡烛在任何观看距离下都有约 1cd 的发光强度。

发光能量的单位是塔尔博特 (talbot) [T]。 1T 是 $1\text{lm} \cdot \text{s}$ 。

在特定方向上的光通量密度是照度 (L_v)。正式地，照度是光通量对立体角和投影面积的微分。照度是辐射率在光度测定的类似量；它以 $\text{cd} \cdot \text{m}^{-2}$ 为单位表示，口语化的单位是尼特 (nit) [nt]。照度是一个重要和有用的量度，因为它在镜头变换下保持不变。 $1\text{cd} \cdot \text{m}^{-2}$ 的照度对应大约每秒每平方度一百万个 560nm 的光子。亮度是照度的感知相关物，但亮度感知是非常复杂和高度非线性的。根据 CIE 的定义，亮度是不可测量的主观量。

一个完美的 Lambertian (漫) 反射体的照度 (cd/m^2 , nt) 是 $1/\pi$ 乘以其光度 (lux)。

B. 3 亮度级示例

表 B-2 给出了日常生活中的亮度级的例子。

表 B-2 亮度级示例

情 景	光度/lx	90% 漫反射体的照度/nt	18% 漫反射体的照度/nt
正午晴朗明亮的日光	100 000	30 000	6 000
典型的日光	10 000	3 000	600
阴天的日光，电视演播室	1 000	30	6
很阴的天空，居室	100	3	0.6
黎明或黄昏，33cm 的烛光	10	0.03	0.006
微光，1m 的烛光	1	0.003	0.000 6
晴空，满月	0.1		
晴空，半月	0.01		
星光+辉光	0.002		
星光，阴天的夜晚	0.000 1		
天狼星 ($m_v = -1.47$)	0.000 01		
织女星 ($m_v = 0$)	0.000 003		

B. 4 影像学

绝对照度以 $\text{cd} \cdot \text{m}^{-2}$ 为单位，俗称 nit [nt]。在影像学中，照度通常相对于一个指定的或隐含的白色参考值归一化到一个 100 的范围内（或用 1 更好），不用单位表示。这样归一化后的符号是 Y。术语照度经常用于这个纯数量的简写表示，但是，它叫做相对照度更合适。应该避免用术语照度因子表示这个量，因为照度因子指的是表面或材料的一种属性：照度因子是在指定的光源、入射和观察条件下，一种表面的照度与一个完美漫反射表面照度的比值。

相对照度 (Y) 是由 CIE 标准化的三个著名的刺激值之一，另两个是 X 和 Z。别的三刺激值，如 $[R, G, B]$ 用一次 3×3 线性矩阵乘法与 CIE $[X, Y, Z]$ 值相关联。相对照度 (Y) 和其他的刺激值，如 X、Z、R、G、B 都是纯数字。

Michael Brill 和 Bob Hunt 赞成 R, G, B 三刺激值不带单位，见 Hunt, R. W. G. (1997), “The heights of the CIE colour-matching functions,” in *Color Research and Application*, 22(5): 337 (Oct)。

B.5 单位

许多奇怪的单位已用于光照度和照度。作者敦促读者放弃这些单位而采用标准的 SI 单位。不用去应付过多的晦涩单位，辐射测定和光度测定已经够困难了。如果读者刚接触辐射测定和光度测定，要是你忽略科学社群 40 年前已抛弃的传统单位，并采用 SI 单位，相信你的理解会更快。如果你是一个用传统的英制单位学习科学和手艺的从业者，请不要顽固地坚持它们：根据美国中央情报局世界概况，只有三个国家，缅甸、利比里亚和美国，没有采纳国际单位制（SI，或公制系统）作为重量和测量的官方系统。作者极力主张你使用 SI 单位。简单学会了用 3.4 乘以英尺·朗伯（fL）得到坎德拉每平方米。电影银幕通常有大约 14fL 白色照度，称其为 48nit。北美的演播室参考显示器通常有大约 100nit（约 33fL）的参考白色照度。

正是美国习惯单位（英制单位！）与 SI 单位不一致的混用导致 1998 年 NASA 的火星气候轨道飞行器坠毁的。美国电影业专家经常用英尺·朗伯表示照度；那他们为什么称电影胶片为 35mm，而不是 $1\frac{3}{8}$ in？

用表 B-3 可将光照度换算成勒克斯（lx）。

表 B-3 将光照度换算成勒克斯

乘以以下单位求得 (lx)	乘以	数值
lm · ft ⁻² ，英尺烛光 (fc)	10.764	10.764
米烛光	1	1

用表 B-4 可将照度换算成坎德拉每平方米。

表 B-4 将照度转换为 cd · m⁻²

乘以以下单位求得 (nit)	乘以	数值
朗伯 (L)	$10\,000/\pi$	3183.1
毫朗伯 (mL)	$10/\pi$	3.1831
cd · ft ⁻²	10.764	10.764
英尺·朗伯 (fL)	$10.764/\pi$	3.4263

延伸阅读

Ian Ashdown 著的书的第一章给出了一个易懂的光测量的介绍（书的其余部分详细介绍了称为辐射度的计算机图形学技术）。网上可以找到该章的一个版本。

D. Allan Roberts 提供了对辐射测定和光度测定基本量的一个精练总结，并描述了那些让人困惑的单位（值得找出 1994 年的版本；不幸的是，2006 年版的手册关于辐射测定的描述谈不上清楚）。

Ashdown, Ian (1994), *Radiosity: A Programmer's Perspective* (New York: Wiley).

Ashdown, Ian (2002), *Photometry and Radiometry-A Tour Guide for Computer Graphics Enthusiasts*, <<http://www.helios32.com/AAsasuring%20Light.pdf>>.

Roberts, D. Allan (1994), “A Guide to Speaking the Language of Radiometry and Photometry,” in *Photonics Design and Applications Handbook*, 1994 edition, vol. 3, pages H-70 to H-73 (Pittsfield, Mass. : Laurin Publications).

本书从基本概念、基础理论、实际问题、演播室标准、视频压缩、发行标准6个方面系统阐述了数字视频与高清技术。作者多年从事数字视频与高清技术方面的学术及工程应用研究，书中对基本概念的诠释极为精准，这在同类书籍中较为罕见。本书适合作为高等院校计算机、电子与信息工程、通信工程、数字媒体等相关专业高年级本科生和研究生的“数字视频技术”课程教材或教学参考书，也可作为工程技术人员的参考资料或培训教材。

本书特色

- 呈现了数字化、采样、量化、伽马、滤波的基本概念。
- 介绍了基于文件的工作流概念，强调了元数据的新兴领域。
- 保留了少量但必要的公式，高质量的图示，便于读者理解理论概念。
- 讨论了新的母带处理中用到的参考显示可观看条件。
- 介绍了立体（3D）视频相关知识。

作者简介

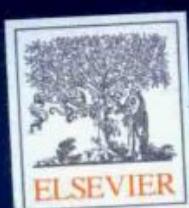
查尔斯·波因顿（Charles Poynton）美国电影与电视工程师学会（Society of Motion Picture and Television Engineers, SMPTE）会士，并于1993年因在计算与通信集成视频技术方面的杰出贡献被授予David Sarnoff金奖。1981年，他创立了伯顿矢量公司，同年为美国宇航局的约翰逊航天中心设计数字电视处理设备。他还著有《A Technical Introduction to Digital Video》。

本书译自原版*Digital Video and HD: Algorithms and Interfaces, Second Edition*并由Elsevier授权出版



投稿热线：(010) 88379604
客服热线：(010) 88378991 88361066
购书热线：(010) 68326294 88379649 68995259

华章网站：www.hzbook.com
网上购书：www.china-pub.com
数字阅读：www.hzmedia.com.cn



上架指导：数字视频

ISBN 978-7-111-56650-2



9 787111 566502 >

定价：99.00元