



OHM 通信实用技术系列

JPEG/MPEG2 技术

〔日〕小野定康 铃木纯司 著
叶 明 译



9.81

86

 科学出版社
www.sciencep.com

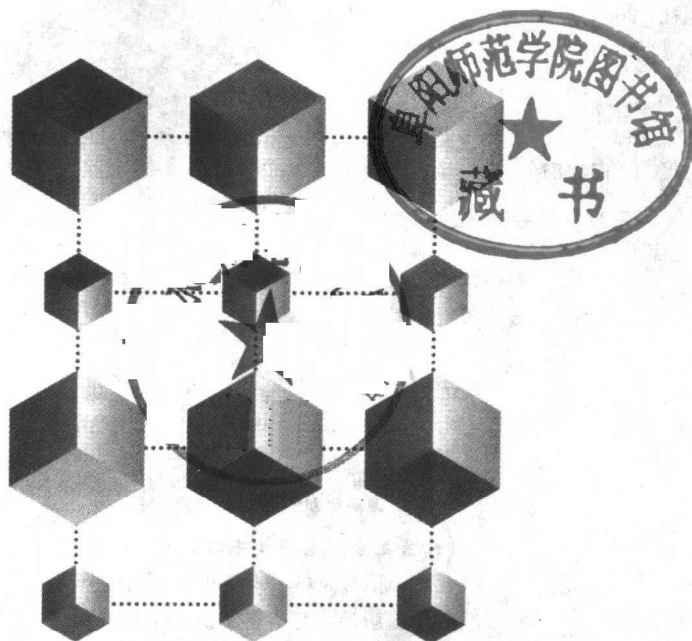
TN919.81

X386

OHM 通信实用技术系列

JPEG/MPEG2 技术

〔日〕小野定康 铃木纯司 著
叶 明 译



科学出版社
北京

图字:01-2003-3495 号

Original Japanese language edition
Wakariyasui JPEG/MPEG2 no Gijutsu
By Sadayasu Ono and Junji Suzuki
Copyright © 2001 by Sadayasu Ono and Junji Suzuki
Published by Ohmsha, Ltd.
This Chinese version published by Science Press, Beijing
Under license from Ohmsha, Ltd.
Copyright © 2003
All rights reserved

**わかりやすい
JPEG/MPEG2の技術**
小野定康 鈴木純司 オーム社 2002

图书在版编目(CIP)数据

JPEG/MPEG2 技术/(日)小野定康, 铃木纯司著; 叶明译. --北京: 科学出版社, 2003

(OHM 通信实用技术系列)

ISBN 7-03-012159-7

I. J… II. ①小…②铃…③叶… III 图像处理 数据压缩 图像编码
IV. TN919.81

中国版本图书馆 CIP 数据核字(2003)第 080387 号

责任编辑 王 炜 崔炳哲 责任制作 魏 瑾
责任印制 刘士平 封面设计 李 祥

科学出版社 出版

北京东城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

源海印刷有限责任公司印刷

北京东方科龙图文有限公司 制作

<http://www.okbook.com.cn>

科学出版社发行 各地新华书店经销

*

2004 年 1 月第 一 版 开本: B5(720×1000)

2004 年 1 月第一次印刷 印张: 13 1/4

印数: 1—5 000 字数: 175 000

定 价: 26.50 元

(如有印装质量问题, 我社负责调换〈新欣〉)

● 前 言

本书主要介绍作为多媒体编码方式的国际标准,即最近几年迅速应用于广播、通信、存储媒体等领域的 JPEG 和 MPEG2。书中前半部分介绍图像压缩算法的有关背景;后半部分对图像压缩算法的内容进行了详尽介绍。

本书的特点是没有只停留在 JPEG 和 MPEG2 的技术性介绍上,而是同时交叉介绍了对其相关的广泛领域的未来发展方向的预测。这是因为目前已经渐渐进入普通家庭的数码照相机、数码摄像机和个人计算机等,在与迅速发展的以因特网为中心的通信系统连接前,我们还不能明确预测到将会打开一个什么样的新世界,而这些正是读者最关心的问题。

因此,本书在前半部分介绍了多媒体系统存在的意义、处理图像信息的性质和削减数据量的基本压缩技术及应用以上技术的各种数字图像系统。对想成为设计多媒体系统的技术者来说,从中可以了解为什么利用 JPEG 和 MPEG2 为多媒体系统的原型,图像质量会达到什么程度,以及在实际的系统中,是以何种方式使用等必须掌握的知识。

为进行多媒体系统的设计,不能只停留在概念的理解上,必须要对 JPEG 和 MPEG2 的标准有一定程度的了解。为掌握最详细而正确的信息,需要仔细阅读这些国际标准方式的正式文件。这需要具有较好的包括数字信号处理和信息理论在内的专业知识。这里所讲的内容不一定要求所有的读者都掌握。但是,如果在多媒体系统设计中,进行算法模拟和高效率的 LSI 开发时,有一本容易正确理解这方面规格本质的参考书会非常方便。为实现这样的目的,在本书后半部分尽力做到对 JPEG 和 MPEG2 算法的讲解详尽而精辟。当然,其中也包含了以这些标准方式为基础进行改良算法研究的研究人员十分需要的信息。

本书是 1995 年日本 OHM 社出版的《易懂 JPEG/MPEG2 的实现方法》一书的修订版。前版承蒙各位读者的支持得以再版,但因为本领

域内技术的发展非常显著,笔者感觉到有些部分已经陈旧。因此,对于前半部分,在更新大量最新信息的同时,也增加了对未来的重新预测。应用 JPEG 和 MPEG2 的多媒体系统就在我们的身边,怎样利用它们来丰富我们的社会生活,已逐渐成为一个重要的课题。本书如果能对担负起这种时代重任的多媒体系统技术者和研究者们有所帮助,我们会感到十分荣幸。

著 者

● 目 录

第 1 章 多媒体系统

1.1 媒 体	2
1.2 多媒体	4
1.2.1 交互式	6
1.2.2 双向性	7
1.2.3 实时性	7
1.2.4 按需应变	8
1.2.5 定制化	8
1.3 协 议	9
1.3.1 TCP/IP	10
1.3.2 应用层协议	12
1.4 硬 件	12
1.4.1 数字(逻辑)线路	12
1.4.2 布线逻辑和程序逻辑	14
1.4.3 PC 机为基础的多媒体系统	14

第 2 章 图像信息及其压缩技术

2.1 图像数据的性质	18
2.2 图像信号的数字化	20
2.3 比特率编码的图像数据压缩	21
2.4 图像品质和内容	23
2.5 进行图像数据压缩的基本技术	24
2.5.1 隔行扫描	25
2.5.2 颜色信号的变换	26
2.5.3 基于视觉特点来抑制高频率部分	27
2.5.4 熵编码	29
2.5.5 运动补偿	29
2.5.6 矢量量化	30
2.5.7 小波变换	30

2.5.8 图像编码算法小结	32
----------------------	----

第3章 数字图像系统

3.1 数字照相机	36
3.1.1 拍摄图像的实时确认	37
3.1.2 数字图像的检索	39
3.1.3 图像的永久性	40
3.2 数字录像机	42
3.2.1 图像劣化	43
3.2.2 完全复制	44
3.2.3 记录媒体的小型化	44
3.3 数字影碟	46
3.3.1 外部记录装置	46
3.3.2 关于随机读取	47
3.3.3 写入的可能性	48
3.3.4 DVD 的图像压缩	48
3.4 卫星播放系统	49
3.4.1 电视播放方式	50
3.4.2 图像品质和显示设备	51
3.4.3 模拟压缩和数字压缩	54
3.4.4 硬件规模 and 成本	55
3.5 PC 用图像卡	56
3.5.1 结 构	56
3.5.2 图像的压缩和解压	57
3.5.3 数字接口	58

第4章 JPEG 的编码算法

4.1 JPEG 概要	62
4.1.1 运行方式和基本构成	62
4.1.2 失真压缩和无失真压缩	63
4.1.3 JPEG 中规定的范围	65
4.2 失真压缩的 DCT 方式	66
4.2.1 利用 DCT 对空间频率的变换	66
4.2.2 利用量化削减等级数	67
4.2.3 熵编码器输入信号的制作	68

4.3	无失真压缩的 Spatial 方式	69
4.4	霍夫曼编码算法	70
4.4.1	编码方法	70
4.4.2	解码方法	73
4.5	算术编码算法	77
4.5.1	二进制算术编码的方法	77
4.5.2	统计编码的方法	80
4.6	JPEG2000	82

第 5 章 JPEG 算法的实现和图像压缩举例

5.1	JPEG 算法的实现	86
5.1.1	变换编码和解码的实现	88
5.1.2	霍夫曼编码解码部分的实现	95
5.1.3	算术编码和解码部分的实现	108
5.2	JPEG 实现的图像压缩	128
5.2.1	基本系统(baseline system)	128
5.2.2	扩展系统	130
5.2.3	可逆方式	130

第 6 章 MPEG2 的编码和算法

6.1	MPEG 概要	134
6.1.1	所谓 MPEG	134
6.1.2	类和等级	135
6.2	MPEG 算法的基本结构	137
6.2.1	无编码器的规定	137
6.2.2	从解码器出发	137
6.2.3	MPEG 编码器、解码器的简易模型	140
6.2.4	MPEG1 和 MPEG2 的相同处理	142
6.2.5	在比较上的各种看法	145
6.2.6	与 MPEG1 的不同点	146
6.3	视频数据的层结构	148
6.4	比特流的结构	151
6.4.1	起始字符	151
6.4.2	比特流的记述方法	152
6.4.3	比特流序列	153

6.5	可伸缩性的导入	157
6.5.1	可伸缩性的概念	157
6.5.2	空间可伸缩性	159
6.5.3	SNR 可伸缩性	159
6.5.4	时间可伸缩性	159
6.6	类和等级的关系	160
 第 7 章 MPEG2 算法的实现和图像压缩举例		
7.1	解码算法的实现	164
7.1.1	DCT 系数的解码	166
7.1.2	逆“之”字形扫描	173
7.1.3	逆量化	174
7.1.4	饱和控制	177
7.1.5	误配控制	178
7.1.6	逆 DCT	179
7.1.7	非编码像块和非编码宏块的处理	180
7.1.8	指定宏块地址和决定像条位置	183
7.1.9	运动补偿	184
7.1.10	宏块类型	194
7.2	MPEG2 实现的图像压缩	196
7.2.1	压缩率和 SNR	198
7.2.2	编码比率的变化	199
7.2.3	图像结构和 SNR	199
参考文献		201

第1章

多媒体系统

JPEG / MPEG2

1.1 媒 体

公元 8 世纪,中国等东亚及东南亚各国已经开始广泛使用木版印刷。到了公元 14 世纪,在朝鲜出现了世界上最初的金属活字印刷。是否使用金属活字是区分木版印刷和金属印刷的主要特征。但是这项发明没有在先前发明的木版印刷基础上普及,并且印刷的对象只限于文字,不能完成图像的一般推广。

从公元 15 世纪,德国人约翰·谷登堡(Johannes Gutenberg, 1397~1468 年)发明了铅活字以来,媒体就逐渐成为推动社会前进的要素。正确地说,与其他一般意义上从无到有的发明相比,只有铅活字版印刷是在漫长的人类历史中,从反复在纸上记录相同图像的无数次尝试中发明出来,并迅速在社会上普及的一种发明。

但是,与其说铅活字印刷术比起以前的各种尝试具有突出的优点,还不如说是社会的需求推动了铅活字印刷术的迅速普及。谷登堡的发明正好处在西欧基督教的天主教(旧教)和新教的传教纷争时期,在这场纷争中印刷起到了重要的作用。

从文字信息向图像信息的发展,无论是印刷还是电子信息都是漫长反复的。但是印刷在向图像信息的发展中,只能停止在静止图像上。而电子信息系统却包含了静止图像和二维运动图像(电视)的发展。与印刷中使用的图像品质相比,电子信息系统在向运动图像发展中存在图像品质低下的问题。因此决定了印刷和电视不能联合,而各自具有不同的适用范围。

以文字为对象的印刷术同以忠实地再现静止图像和容易复制为目标的照片连接在一起,又经过了相当时间^[45, 46]。照相技术是 1816 年由法国人约瑟夫·尼埃普斯(Joseph Niépce, 1765~1833 年)发明的,然后经过法国的路易斯·达盖尔(Louis Daguerre, 1787~1851 年)、英国的威廉·亨利·福克斯·塔尔博特(William Henry Fox Talbot, 1800~1877 年)、法国的路易·迪克·迪·奥隆(Louis Ducas Du Hauron, 1837~1920 年)及美国的乔治·伊斯曼(George

Eastman, 1854~1932 年)等人的努力,确立了其作为完备技术的牢固地位。但是,照相不只是以化学胶片为基础的静止图像,法国的卢米埃尔(Lumière)兄弟[奥古斯特(Auguste), 1862~1954 年;路易(Louis), 1864~1948 年]和美国人托马斯·阿尔瓦·爱迪生(Thomas Alva Edison, 1847~1931 年)将其发展到被称为电影的一个新的运动图像领域。因为这种电影的图像品质高于电视,所以至今电视还在寻求使用胶片来实现电影的图像品质。关于照片和电影的发展年表如表 1.1 所示。

表 1.1 照片和电影的发展年表

1729 年	舒尔策(Schulze),发现氯化银在阳光的照射下会发生变化
1816 年	尼埃普斯,在涂有氯化银的纸上成功得到第一个曝光图像(负片)
1826 年	尼埃普斯,发明使曝光图像定影的“阳光摄影法”(heliography)
1834 年	塔尔博特,开发在感光纸上定影图像的方法,成为负片与正片冲印技术的基础
1839 年	达盖尔,公开“银版照相法”(daguerreotype)
1855 年	迪·奥隆,研究连续摄影
1878 年	穆布里奇,成功地连续拍摄了一匹奔跑的马
1884 年	伊斯曼,发明由赛璐珞制作的胶卷(1889 年发售)
1888 年	爱迪生、伊斯曼,使用 35mm 胶卷完成第一次动画摄影试验
1893 年	爱迪生,发明电影视镜的原型“窥见型放影机”(kinetoscope)
1895 年	卢米埃尔兄弟,使用“活动电影机”(cinématographe),将动画投射到银幕上

因为照相和印刷都是以容易大量复制为目的的,所以它们很快就合为一体。这种印刷和照相结合确立的照相制版(胶印)经历了相当长的一段时间。现在,胶印代替了谷登堡的铅活字版印刷成为印刷业中的主角,这是一个具有深远意义的事实。现实是,谷登堡苦心发明的铅活字版印刷在现在的印刷业中已经难寻踪迹了。

随着照相制版的完全电子化,胶印也发生着巨大的变化。现在广泛使用的个人计算机(PC:Personal Computer)用打印机作为胶印媒体的主角地位也已经开始动摇。一种可以廉价地大量即时发送和现有的印刷物具有同等的易读性的高

品质图像的随选自主印刷(on demand-personal printing),正在改变着拥有多媒体的人群。可以预测,其结果将可与谷登堡的铅活字版印刷术带来的社会结构变化相匹敌。

本书主要介绍作为印刷-照片技术的延续发展的电子图像技术中的静止图像和运动图像的比特率编码/压缩(bit rate reduction/compression)算法。众所周知,这些算法已经成为世界标准,并被广泛应用到各种领域中。

1.2 多媒体

“多媒体”一词由来已久。多媒体是抽象的,因而不能简单地说是“这个”就是多媒体”。

若根据事实来定义多媒体的话,则如图 1.1 所示,它是可以同时提供声音和图像的系统。即要满足以下五条:

- ① 交互式
- ② 双向性
- ③ 实时性
- ④ 按需应变(on demand)
- ⑤ 定制化

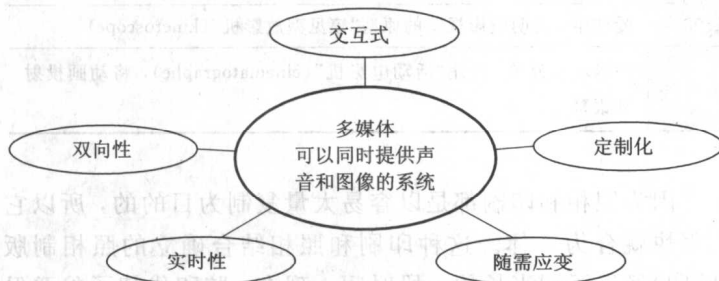


图 1.1 多媒体的定义

但是,这些条件之间具有相互关系,也可能有不满足其中一部分或不能完全满足该部分的情况。

尽管“多媒体”一词的频繁使用已经是十几年的事情了,但却几乎没有被大众所认知的具体的多媒体系统,这不能不说是一种怪现象。当然也不能断然地判断不存在多媒体系

统。例如,日本生产的游戏机就可以完全满足多媒体的定义。但是把游戏机作为多媒体的代表,明确地把多媒体和游戏机联系起来的恐怕只有笔者。当然,在这方面的理解上,人与人之间是存在差异的。例如,一说起“广播”人们就会想起电视、收音机,一说起“出版”人们就会想起杂志、图书。而提起“多媒体”一词就会感到缺乏具体感,但笔者认为这不仅仅是因为该词的使用时间较短所造成的。

理由大致可以分为两方面:其一,“多媒体”一词是受到了已有媒体或只有声音,或只有图像,被动的、单方向的等“框框”的制约,而对这个“框框”以外的东西就只具有模糊的期待,也就是说,是作为缺乏现实性的“青鸟”¹⁾被使用着。

其次,无论是日语还是英语中,“媒体(media)”一词的意思都是非常宽泛的,换言之,是被模糊地使用着,因而从模糊中产生了缺乏实用性的缺点。“media”一词起源于法语中表示“中间的”意思的“medium”的复数形式。总之,是泛指人与人之间传递信息所使用的工具。因此,如果把保存信息的媒介 MO(Magneto-Optics)盘也称为媒体的话,那么广播电视局、出版社等组织也可以称为媒体。像这样意思上的宽泛性,即不能明确指出对象的模糊性,经常影响着“媒体”一词。

今后无论多媒体怎样变化发展,多媒体系统都不能改变它传送声音和图像信息的功能。而为了满足前述的多媒体的五点特性,能正确控制声音、图像信息及高速的通信线路是不可缺少的。所谓正确的控制可以理解为正确的编辑和加工,而事实上,如果不把声音和图像信息作为数字化信号来处理是不可能实现的。并且要对这些声音和图像进行正确的编辑和加工,需要能自由对声音和图像进行操作的计算机。因此,多媒体系统经常与 PC 机和数字信号处理联系到一起。数字信号处理是在 PC 机上执行还是在其他硬件上执行,是要看该次加工需要怎样复杂的处理和是否需要实时处理结果等因素,具体情况具体分析,不能一概而论。

当今,PC 机通常与 10Mbit/s 以上以太网(Ethernet)这样的高速通信线路相联接。而这对 PC 机来说是将文字、声

1) 比利时作家莫里斯·梅德朗克(Maurice Maetrlinck)写的童话《L'Oiseau Bleu(青鸟)》,富含深邃哲学思考,在日本广为人知。——译者注

音和图像信息与其他 PC 机共享(包含传送)所不可缺少的条件。高速通信线路是短时间或者实时传送大容量的数字化声音和图像信息所必需的。

从这种角度看,多媒体系统是以 PC 机、高速通信端口和 JPEG/MPEG 为主要结构要素的系统。即多媒体系统从硬件上来说就是这样构成的。关于这些将在 1.4 节中再次介绍,但这套硬件系统将被如何称呼还取决于今后的利用形态。

1.2.1 交互式

“交互(interactive)”原本是相互作用的意思。一般在与计算机相关的范围中是指用人机对话的形式来进行操作。换言之,它是指命令计算机执行某种操作时,计算机和用户界面之间通过按部就班地询问必要项目来进行的形式。

这些项目分为必须指定的项目和不一定要指定的项目。当然如果指定项目,会在最理想的形态下执行命令。在可以不指定的项目中一定会存在默认的解释,如果没有指定会使用按照默认解释预设的值,即默认值(default value)。在没有做这种默认解释的系统中,交互有时会变得非常繁琐。

精炼并容易使用的交互式用户界面要善于利用类推,隐喻(暗示)在这方面起到了十分重要的作用。但是隐喻经常受用户所具有的知识的不明确性的影响,因此,把这种不明确性在某种程度上进行有效地排除的一贯设计方针的存在具有相当大的意义。

解决这种问题需要获取用户在使用体验中的反馈信息。并且这种操作的交换不能只使用文字,使用被称为图标(视觉象征)的图形也是提高用户界面性能的一种有效的方法。但是,因为图形的本身具有多义性,所以在使用时要充分考虑。进一步,在思考中“隐喻”完成了根本的任务,有必要确认它与视觉的结合。总的来说,思考和图形操作是不可分割的,文字是最单纯的抽象化图形,所以其不确定性最小。但是只对该部分进行操作时,需要训练和预备知识。如果将对 PC 机进行命令操作与进行图标操作相比较,这件事情就很容易理解了。

1.2.2 双向性

人与人的交流是相互之间传递和获取信息的过程,这是非常自然的。我们把这种相互交换信息称为双向性。不是说什么样的交流都需要具有双向性,但最好在需要的情况下能确保交流的双向性。

电视、广播和传真都不具有通信的双向性,只是单方向的传递信息。而电话、可视电话和电视会议可以进行双向的信息传递。是否需要双方向通信取决于进行通信的目的。以因特网为基础的电子邮件是单方向传递的,但同样在因特网上的 WWW(World Wide Web),因为是交互性的操作,所以决定了它是双向性的操作。

1.2.3 实时性

传送的信息在对方没有明确意识到延迟的时间内传到,这种情况被称为实时性,或即时性。一般地,在双向性通信中,因为要立即得到对方的回答,所以需要该延迟要很短。但允许的延迟时间的最大值是很难正确描述的,通常情况下该值不能超过 $2\sim 3s$ 。

产生延迟时间的原因可以分为物理因素和方式因素。物理因素是因为光、电波、电流具有传输速率,而它们是不能缩短的;方式方面是指内存的读写速度、缓存器的容量的原因,如果不考虑经济因素,可以大量缩短延迟时间。并且随着技术的进步,延迟时间也可以得到进一步的改善。但是,由物理因素造成的延迟问题,通常是超过 1000km 的超长距离国际通信造成的。

以因特网为基础的电子邮件,是不具有实时性的存储和转送(store and forward)通信方式,何时到达对方是由网络的状况来决定的,而 WWW 是交互式的操作,希望在最短的等待时间内得到操作结果,因此要求具有很强的实时性。

1.2.4 按需应变

在需要时可以直接使用,这种情况称为按需应变(on demand)。如果确保该条件和实时性,就可以实现无需等待直接返回对方的回答,极大方便用户。但这同时引发了一个非常庞大的经济方面的问题。它是由在实际中使用的不确定性而引起的各种通信资源待机所造成的。换言之,不在一定程度上抑制通信资源的共用性就不能实现按需应变。当然在不同的情况下,也存在完全不可以共用的资源。

在通信方面,专线(leased line)就是为适应这种按需应变的要求而准备的。如果控制通信资源的共用性,自然会造经济方面的问题。一般通信具有统计方面的特性,从这种统计方面的特性来看,很多情况下某些程度的通信资源可以共用。但是,因为利用这种统计方面的特性不可能实现完全的按需应变要求,所以评价按需应变要求的强度是通信系统中不可欠缺的事项。

1.2.5 定制化

一般地,如果机器功能复杂,人们希望实现某种功能的必要操作最好是符合自己习惯的形式。只有一种方法实现某种功能的操作,对某些人来说是简单的,但对其他人可能是繁琐的,因为没有能够同时都满足的方法,所以预置能适合各种人群的操作方法,成为惟一的解决方案。很多情况下,不可能有同时满足初学者和有经验者的操作方法。我们把为解决这种问题而采用特别定做以适应特定个人的操作方法称为定制化。

在PC机中基本的图标操作方法是适用于初学者的,但这种操作对有经验者来说是比较繁琐的,因此准备了称为快捷方式的操作量小的操作方法。如果用户可以自己定义适宜的操作方法,那就更加方便了。通常这样的操作是恰当地使用了UNIX中(command)的外壳(shell)。而与此相对应的是内核,因为是操作系统(OS: Operating System)的核心组成部分所以称为内核(kernel)。对于定制化中应该注意的