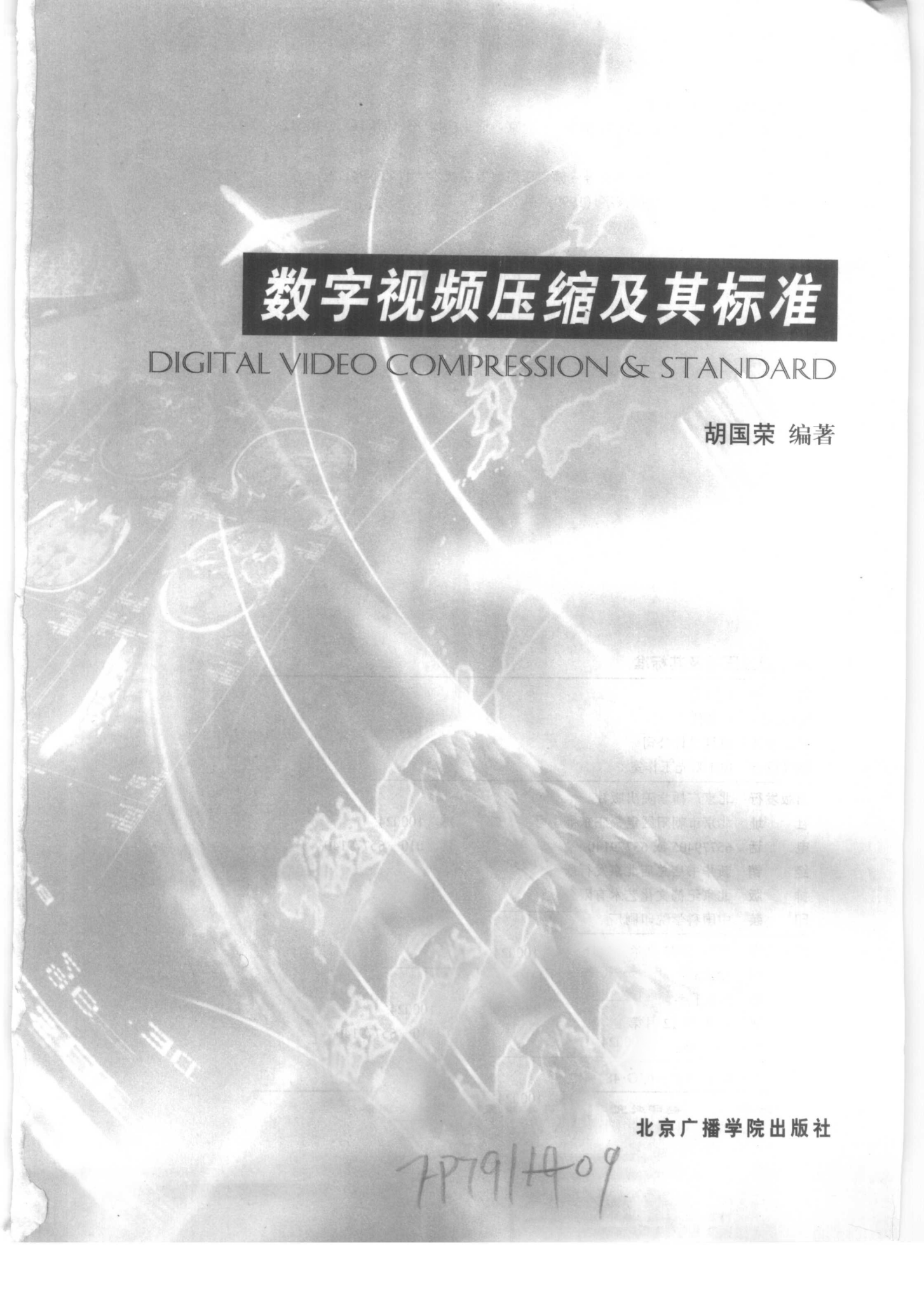




数字视频压缩及其标准

DIGITAL VIDEO COMPRESSION & STANDARD

胡国荣 编著

北京广播学院出版社

7P79/H09

图书在版编目 (CIP) 数据

数字视频压缩及其标准/胡国荣编著. - 北京: 北京广播学院出版社, 1999.12

ISBN 7-81004-837-6

I. 数 II. 胡… III. 视频信号-频率压缩-数字技术 IV. TN941.3

中国版本图书馆 CIP 数据核字 (1999) 第 75267 号

数字视频压缩及其标准

编 著 胡国荣

责任编辑 叶桂刚

封面设计 恒真设计公司

版式设计 BBI 阳光工作室

出版发行 北京广播学院出版社

社 址 北京市朝阳区定福庄东街 1 号 邮 编 100024

电 话 65779405 或 65779140 传 真 010-65779140

经 销 新华书店总店北京发行所

排 版 北京纪德文化艺术有限公司

印 装 中国科学院印刷厂

开 本 787×1092 毫米 1/16

印 张 12.75

字 数 290 千字

版 次 1999 年 12 月第 1 版 1999 年 12 月第 1 次印刷

印 数: 1—3000

ISBN 7-81004-837-6/G·484

定 价 23.00 元

版权所有

翻印必究

印装错误

负责调换

怒发冲冠，
凭阑处、潇潇雨歇。
抬望眼，
仰天长啸，壮怀激烈。
三十功名尘与土，
八千里路云和月。
莫等闲，
白了少年头，空悲切。

与同学们共勉

Preface

Chiariglione Leonardo
(MPEG Chairman)

In the late 1980s a number of initiatives were dealing with the problem of digital coding of audio and video: the CCITT H.261 recommendation, the DVI (digital video interactive) of the David Sarnoff Research Center, FMV (full – motion video) for CD – I (compact disc interactive) of Philips, the 34/45 Mbit/s contribution codec of the CMTT, the ATV initiative of the FCC (Federal Communications Commission) in the USA, the HDTV codec for satellite broadcasting of RAI and Telettra, the DAB (digital audio broadcasting) systems developed by the European project EU 147.

Today it may seem preposterous that in 1988 one could have established yet another and brand new group at the experts level, the Moving Picture Experts Group (MPEG) in the framework of a virtually unknown ISO committee (TC 97/SC 2 "Character Sets and Information Coding") to achieve something apparently similar to what other groups with an established tradition were already doing. And so it seemed to be at the first meeting in May 1988 when 15 participants in a single session attended.

Eleven years after the landscape of audio and video has completely changed: several tens of million Video CD players have already been sold, Windows has an MPEG – 1 software decoder, MP3 – MPEG – 1 Audio Layer III has taken the music world by storm, several tens million digital television set top boxes for satellite and cable have been sold, several millions DVD players have been sold; digital terrestrial television services have started in several countries.

There were differences compared with those other initiatives, and these were the principles on which MPEG had been established: the necessity to deal jointly with audio and video, in spite of the organisational barriers that traditionally existed between these two technical communities, the identification of digital audio and video as a generic technology to serve the needs of multiple industries because a common standard would create mass of cross – industry applications based on the common technology and the need to get the involvement of the different industries to maintain a high degree of interoperability across applications.

MPEG is a success story. It has used the know - how developed by the different components of the multimedia industry and has integrated it in powerful standards that have responded to the needs of the times. MPEG could do this because it was able to convert business requirements into technical specifications for the work, while putting a firm dividing line between the two. Then it set out developing its standards by exploiting the enthusiasm of hundreds of researcher from all parts of the world, unencumbered by immediate business considerations.

•
May. 13, 1999

Translation in Chinese (中文译文)

序 言

Chiariglione Leonardo (MPEG 主席)

在二十世纪八十年代末期, (国际上) 出现了一批解决数字音频和数字视频问题的倡议, 包括: CCITT 的 H.261 建议, David Sarnoff 研究中心的 DVI (交互式数字视频) 方案, Philips 的 CD-I (交互式光盘) 方案, CMTT 的 34/45 Mbit/s 编解码方案, 美国 FCC (联邦通信委员会) 的 ATV (先进电视) 倡议, RAI 和 Telettra 的卫星广播 HDTV (高清晰度电视) 编解码方案, 欧州 EU147 计划的 DAB (数字音频广播) 系统方案。

今天看来不可思议的是: 在 1988 年, 某人 (即 Leonardo 先生本人) 居然能标新立异开创一个新的专家级组织 - MPEG (运动图象专家组) 并在一个不为广知的国际标准 (ISO) 委员会下属组织 - "TC 97/SC2" (符号集与信息编码) 里从事明显类似与其他组织按惯例已经进行的工作。在 1988 年 5 月第一次会议召开时只有 15 个与会者参加了唯一一个会议。

十一年过后的今天, 音频和视频领域已经完全改天换地: 成千上万的 VCD 机已经被销售, Windows 已经有 MPEG-1 的软解压器, MP3 (MPEG-1 音频第三级标准) 的音乐象潮水一样涌入世界, 数以百万的卫星和有线电视的数字电视机顶盒被销售, 无数 DVD 播放机被销售; 数字地面电视服务已经在几个国家开始。

与其他的倡议相比, MPEG 有很大不同, 正如 MPEG 专家们建立的原则: 必须把音频和视频结合在一起 (尽管在这两个技术领域传统上存在组织上的障碍) 以使得数字音频和数字视频成为统一的技术来满足各种不同的工业领域, 因为一个共同的标准可以在共同的技术基础上产生一大批交叉工业并保持不同的应用领域的高度兼容性。

MPEG 是一个成功的故事。她紧随时代的要求利用多媒体工业各方面开发的技术并集成到强有力的标准中。MPEG 之所以能做到这些是因为她能够把市场需求转化为技术指标, 同时在两者之间划上严格的界线。然后开发世界各地无数研究者的热情来制定标准, 而不受即刻市场的妨碍。

1999 年 5 月 13 日

前 言

世界正迈进数字化、网络化、全球一体化的信息时代。人类的生活也将在这个变为“地球村”的星球上进入更高品质的“数字化生存”。视觉信息作为人类最“赏心悦目”的信息将大放异彩。而在全球范围内无远弗届实时或非实时地高质量传播视频信息将很大程度上依赖于数字视频压缩及其标准。

众所周知，人类对微观世界的认识已触及基本粒子的结构，对宏观世界的探索已超出了本银河系，对生命本身的研究也到了DNA分子水平，而对视觉的研究也几近“色即是空”的境界。视频图象无论是“急流飞瀑”，还是“行云流水”，都可看作在视觉时间灵敏度范围内对视频信号进行时间采样的一幅幅静止图象组成。而一幅静止图象无论是“闭月羞花”，还是“沉鱼落雁”，都可看作在视觉空间灵敏度范围内对图象进行空间采样的一个个像素组成。然而一帧1024X768个像素的原始图象，如果每个像素平均用12比特（亮度8比特，色度4比特）来表示，则总共需9M bit。按30帧/秒的视频速率，则在1秒钟内数字视频的数据量是270M bit。1分钟之内的电视/电影数据就将占满现有高档PC机的所有硬盘和内存。如果不用压缩技术而用现有媒质和信道传播实用的视频图象序列，纵有“倒拔杨柳”的力气和“还看今朝”的气魄也要发出“蜀道难”的叹惜。因此图象/视频信号的压缩是视觉信息传播的至关重要的技术。

一般而言，静止图象具有空间连续性，运动图象（视频信号）除具有空间连续性外（帧内），还具有时间连续性（帧间）。例如一段“落霞与孤鹜齐飞，秋水共长天一色”的视频信号，其中的“落霞”“秋水”“长天”在图象的大部分面积（空间）内是相似的或连续相关的，而“孤鹜”是从前一帧图象中“飞”过来的，其本身和背景都具有时间连续性。图象/视频压缩即是在时域、空域及频域（变换域）描述图象的连续性（相关性）并根据人类视觉特点，去掉空间相关性和时间相关性（信息冗余），使大部分数据都变为0，从而用较少的有值数据有效地表示图象/视频，进而利用信息编码理论（熵编码）对这些0值数据和有值数据进行的数据压缩。其逆过程（解压缩）可以将图象重建。现代高速高数据吞吐率的数字计算技术有力地支持了这些压缩与解压缩处理，使得一幅诸如“波光里的艳影”的高清晰度的图象/视频可以在片刻间“揉碎”和“重现”，使之“轻轻的走”又“轻轻的来”。

近年来，由于视频市场的大力牵引和电子技术的飞速发展，各种各样图象压缩和数据压缩的算法应运而生，层出不穷。为了视频信息及其产品可在全球交流和使用，有必要对这些算法及视频数据格式进行统一并制定相应的国际标准。各国和各大有关跨国公司都极其重视这些标准的制定并都试图使自己的技术专利在国际标准中占一席之地。而众多的技术专利要经过国际专家组进行会议评比和筛选。尽管各国视频界的精英们在制定标准的过程中各竞风流，但每次评选无异是国际视频界的一次“奥林匹克”赛。竞争是残酷的，入围者可坐吃专利，后福无边；淘汰者前功尽弃，不知何处是岸。但由于大家都遵从相同的

“游戏规则”，在“公开、公平、公正”的基础上竞争，赛前没有“高低贵贱”之分，也没有“舍我其谁”的豪气。赛后的胜利者自然令人“刮目相看”，“敬如上宾”。选择虽然有国际专家组人为的因素，但同样是“物竞天择”的结果。

国际运动图象编码专家组（MPEG）成立于1988年。它是ISO/IEC下负责开发运动图象、音频及其组合的压缩、解压缩、处理和编码的国际标准的一个工作组（WG11）。其成员分布在全世界范围内（300-400个专家）的各种研究机构和公司。其工作方式是每年开3-4次国际会议（每次会议一般为一个星期），会议期间分成各种不同的特设小组（AD Hoc Group: AHG）进行讨论，重大提案或表决以“国家队”（National Body: NB）的形式进行。会前所有要讨论的文件都以电子文件的方式上传（Upload）到MPEG公用的FTP站点以便其它MPEG组员下载（所以这种国际会议没有纸质印刷的会议录，一切文件都以电子出版的方式高效地进行）。会后的工作以电子通信的方式进行，由AHG主席进行协调。从1988年到1992年和1990年到1994年间，MPEG分别成功开发了获EMMY奖的MPEG1和MPEG2：

- MPEG-1，视频及其伴音存储标准（ISO/IEC 11172），其码率约1.5Mb/s
- MPEG-2，数字电视标准（ISO/IEC 13818），其码率可达10Mb/s。

这2个标准的关键技术都采用了运动补偿（去时间相关性）、DCT变换（去空间相关性）和Huffman编码（去数据相关性）。不过MPEG2在MPEG1的基础上增加了隔行扫描、伸缩编码和抗错能力等功能。这2个标准极大地推动了相关产业的发展。MPEG-1已经用于成千上万的VCD。事实上INTERNET网上的大部分视音频数据也是用MPEG1方式存储的。DAB也将是MPEG1音频的又一大新的消费市场。而MPEG-2对于模拟电视到数字电视的转换起到了举足轻重的作用。新一代的高清晰度电视也将基于MPEG2标准。无论是卫星广播或电缆广播，成百万的MPEG-2解码机顶盒都在最近3年内“飞入寻常百姓家”。

预见消费电子（电视）、通信和计算机（3C）三大产业将在数字化基础上迅速汇聚，MPEG从1993年7月开始开发针对多媒体应用于多领域的标准，也是其第三个标准—MPEG4。最近（1999年3月）宣布已经完成了MPEG-4第一版本的开发，1999年将成为国际标准ISO/IEC14496。

MPEG-4视频标准集近年来图象分析，图象压缩，视频压缩，计算机视觉，信号处理等领域的最新研究成果的大成（参加竞争的各种技术方案来自并不局限于MPEG的遍布世界各地的各种企业和国家的研究机构），在推出“音视频元”（audiovisual）概念的基础上提出基于具体内容（Content-based）的视觉目标（Visual Object）的编码标准。在开发低码率（5-64Kb/s）编码标准的同时，将重点放在人们更感兴趣的图象具体目标的交互性和可操作性上，并对多媒体多应用领域的编码进行兼容并包。它不仅包括运动目标的编码，还包括静止目标和计算机人工合成目标的编码。它不但是第一个让用户在接收端对画面可进行操作和交互式访问的标准，而且由于其低码率，高度灵活性，兼容性，可伸缩性，强抗错性和可扩展性，也将是第一个解决信息产业中电视、通信、计算机这三大支柱产业的所谓“数字汇聚”问题的标准。

由于MPEG4集中了当代图象/视频压缩技术的精华，本书将着重基于MPEG-4视频标准进行阐述。象MPEG1/2一样，MPEG-4标准其本身只定义其码流语法和解码过程，

对编码算法和过程并不做规定和描述。为了让读者更为清晰地了解 MPEG-4 视频压缩及其标准,本书对其编码和解码都将做一些介绍。

本书首先概述 MPEG-4 视频标准及视频的数字化与码流,然后逐个介绍 MPEG4 形状编码、运动估计和补偿、纹理编码、伸缩编码、灵影编码、差错回避和静止图象编码,最后介绍 MPEG-4 视频标准的应用尤其是在广播领域的应用。附录里搜集了 MPEG4 码流语法全集以及编者在 MPEG4 会议上的发表的文章。本书主要是为北京广播学院的本科和研究生的教学而编写的。与介绍作者本人研究成果的书不同,本书只是介绍作者在实现这些技术的过程中对他人成果的理解和概括。但愿本书能成为同学们进入数字视频领域的一块“敲门砖”。由于编者学识所限,错误难免。书中的一些专有名词翻译如“Sprite-灵影”“Scalability-伸缩性”由于没有参考标准属编者“即兴之作”,有待规范。欢迎商榷和提出意见,以共同提高,使将来的修订版日臻完善。

本书的编写自始至终都得到了北京广播学院院长刘继南教授、副院长苏志武研究员的鼓励和关心。如果没有二位的远见卓识和大力支持,笔者不可能有机会接触和投入到这些最先进技术的领域,也不可能编出这本书。

笔者非常荣幸的是, MPEG 主席 Chiariglione Leonardo 先生在百忙之中为本书做序。在此深表谢意。

笔者在此还要特别感谢 MPEG4 Video AHG 主席 Panasonic Singapore Laboratories 的 T.K.Tan 博士,书中的很多内容都是我们一起工作时的讨论心得。

At last, but not the least, I should thank my wife, Tian Yu and all those who made this book possible.

胡国荣

1999 春于沁园

目 录

序言	(1)
前言	(1)
第一章 MPEG-4 视频标准概述	(1)
1.1 场景的多媒体目标描述	(1)
1.2 视频目标编码	(3)
1.3 基于 VOP 的编码	(6)
附 Summary of MPEG-4 Video Coding and Bitstream	(8)
第二章 视频的数字化与码流	(15)
2.1 灰度图像的数字化表示	(15)
2.2 彩色图像的数字化表示	(16)
2.3 视频图像的采样格式	(17)
2.4 MPEG-4 视频码流结构	(19)
第三章 形状编码	(24)
3.1 二值 alpha 平面	(24)
3.2 BAB 的编码	(25)
3.2.1 编码方式选择	(25)
3.2.1.1 BAB 可接受质量 (ACQ)	(25)
3.2.1.2 编码方式选择	(25)
3.2.2 运动估计 (ME) 和补偿 (MC)	(27)
3.2.2.1 运动矢量 (MV) 的预测	(27)
3.2.2.2 MVs 的计算	(27)
3.2.3 码率控制	(28)
3.2.3.1 VOP 级大小转换	(28)
3.2.3.2 BAB 级大小转换	(28)
3.2.4 基于 CAE 算法的 BAB 编码	(29)
3.2.4.1 算术编码	(31)
3.2.4.2 INTRA and INTER CAE 的邻近信息	(33)
3.2.4.3 BAB 的外围边界信息	(34)
3.2.4.4 基于邻近信息的形状算术编码	(34)
第四章 运动估计和补偿	(46)
4.1 填补过程	(46)
4.1.1 水平填补	(46)
4.1.2 垂直填补	(48)

4.1.3 扩张填补	(48)
4.2 运动估计	(49)
4.2.1 改进的块匹配——条块匹配	(49)
4.2.2 搜索算法	(50)
4.2.3 帧内/帧间预测编码方式选择	(50)
4.2.4 16×16 块和 8×8 块预测方式选择	(50)
4.2.5 半采样搜索	(51)
4.3 运动矢量的差分编码	(52)
4.3.1 基于 16×16 宏块运动补偿的运动矢量编码	(52)
4.3.2 基于 8×8 块式运动补偿的运动矢量编码	(54)
4.4 越过 VOP 边界的运动估计/补偿 (Unrestricted Motion Estimation/Compensation)	(55)
4.5 重叠运动补偿	(56)
4.6 双向运动预测	(57)
第五章 纹理编码	(59)
5.1 8×8 子块二维 DCT 变换	(59)
5.1.1 DCT 变换	(59)
5.1.2 LEP 填补技术	(60)
5.1.3 自适应帧/场 DCT	(61)
5.2 码率控制和量化步长选择	(61)
5.3 量化	(64)
5.3.1 Intra 方式直流分量 (DC) 的非线性量化	(64)
5.3.2 交流分量 (AC) 的量化	(65)
5.4 Intra-DC 和 Intra-AC 的帧内预测	(66)
5.4.1 预测方向	(66)
5.4.2 DC 的自适应预测	(67)
5.4.3 AC 的自适应预测	(67)
5.4.4 AC 预测选择	(68)
5.5 熵编码	(68)
5.5.1 一维化扫描	(69)
5.5.2 基于 Huffman 编码的 VLC	(69)
5.5.2.1 Intra-DC 的编码	(69)
5.5.2.2 其他系数的 Huffman 编码	(70)
5.5.3 逃逸码	(70)
5.6 纹理解码和 MPEG4 视频解码	(70)
第六章 伸屈编码	(72)
6.1 空间伸屈编码	(72)
6.1.1 降采样 (Downsampling)	(73)
6.1.2 基本层的编码	(73)

6.1.3 增强层的编码	(73)
6.1.3.1 升采样	(73)
6.1.3.2 增强层编码	(73)
6.1.3.3 P-VOP 式增强层编码	(73)
6.1.3.4 B-VOP 式增强层编码	(74)
6.2 时间伸屈编码	(74)
第七章 灵影编码	(77)
7.1 简介	(77)
7.2 全局运动补偿	(78)
7.2.1 几何变换	(78)
7.2.2 参考点的定位	(79)
7.2.3 变换函数的定义	(80)
7.3 灵影的产生	(82)
7.3.1 透视运动估计	(82)
7.3.1.1 3 步式匹配	(82)
7.3.1.2 梯度下降算法	(82)
7.3.2 利用透视运动估计的灵影产生	(83)
7.4 灵影编码	(83)
7.4.1 灵影的传输	(84)
7.4.2 灵影形状和纹理编码	(84)
7.4.3 运动轨迹编码	(85)
7.4.4 灵影卷曲	(85)
7.5 灵影解码	(86)
第八章 差错回避	(87)
8.1 重同步	(87)
8.2 数据恢复	(87)
8.3 错误隐匿 (error concealment)	(88)
8.4 错误回避处理	(89)
第九章 静止图像编码	(90)
9.1 基本编码原则	(90)
9.2 离散小波变换	(92)
9.3 对称性扩展 (Symmetric extension)	(95)
9.4 直流子带 (DC sub-band) 的编解码	(98)
9.5 交流子带的量化	(98)
9.6 交流子带的 ZT 扫描	(99)
9.7 熵编码 (Entropy Coding)	(102)
9.8 静止图像的解码	(107)
第十章 MPEG-4 视频标准的应用	(108)
10.1 MPEG-4 的档次和级别 (Profile 和 Level)	(108)

10.2	MPEG-4 应用分类	(111)
10.3	MPEG-4 应用实例	(113)
10.3.1	实时通信	(113)
10.3.2	移动多媒体	(114)
10.3.3	DVD	(114)
10.3.4	基于 Internet/Intranet 的码流视频	(115)
10.3.5	广播	(116)
10.3.6	数字电视机顶盒	(118)
参考文献		(120)
附录 1	MPEG4 码流语法	(122)
附录 2	MPEG4 Dublin Meeting Contribution (M3623)	(176)
附录 3	MPEG4 Atlantic Meeting Contribution (M4062)	(181)

第一章 MPEG-4 视频标准概述

MPEG-4 视频标准在多媒体环境下提供一个基于不同目标的视频描述方法, 包括自然或人工合成视觉目标 (Visual Object) 的压缩、目标伸屈、时空伸屈、差错回避的算法和工具的核技术以有效地用不同媒质存储, 通过现有和将来的有线和无线通信网、INTERNET 网和广播频道传输和操作视频数据。

MPEG-4 视频标准提供一整套技术标准以满足多媒体作者, 网络服务商和最终用户的要求:

- 对于多媒体作者, MPEG-4 将可以使基于数字电视、动画、网页等内容的制作具有极大的灵活性和可重复利用性。并易于保护和产权管理;
- 对于网络服务商, MPEG4 将提供各网络之间的交流信号, 这些信号将有助于各向异性的不同网络之间传输最优化;
- 对于最终用户, MPEG-4 将在小型和较大型终端上提供诸如实时通信、警戒和移动多媒体的很多交互性功能。

与 MPEG1/2 基于帧 (Frame-based) 的压缩标准不同, MPEG-4 是基于目标 (Object-based) 的压缩标准。所以首先它在系统级要定义基于目标的场景描述方法。

1.1 场景的多媒体目标描述

MPEG4 在编码前首先要对视频序列进行分析和理解以提取目标。其码流信息首先给出各个目标的场景描述。想象一幅“人在旅途”的场景:

“枯藤老树昏鸦, 小桥流水人家, 古道西风瘦马, 夕阳西下, 断肠人在天涯。”

其场景可分解 (或分割) 成多个多媒体目标。其原始目标包括:

- 静止图象。如固定的背景“枯藤”“老树”, “小桥”, “人家”, “古道”, “天涯”
- 视频目标 (VO: Video Object)。如“昏鸦”, “瘦马”, “夕阳”, “人”
- 音频目标。如“昏鸦”呜呜, “流水”潺潺, “西风”唳唳, “瘦马”长嘶, “人在”短叹;
- 等等。

一幅复杂的画面就由这些可操作的原始目标组成。如果对这些目标分别进行编码, 最终用户便可以自由地操纵这些原始目标 (如目标的坐标, 视点, 动画等), 还可得到一些原始目标的信息。如对“断肠人”愿观其详以评头品足, 可在观看场景的同时用鼠标点击“断肠人”也许能得到此人的各种信息或网页, 譬如此人名叫“马致远”、“苏东坡”还是“柳永”; 爱好是“声色犬马”、“名山大川”还是“浪迹江湖”; 人品是“孤傲不群”、“气吞山河”还是“从善如流”; 身材是“伟岸高大”、“小巧玲珑”还是“不偏不倚”; 长相是“貌若惊鸿”、“摄人魂魄”还是“不敢逼视”; 擅长是“琴棋书画”、“百步穿杨”还是“治

国平天下”等。

当然计算机对场景的分割只能“见山是山，见水是水”，对其苍凉的意境要做“见山不是山，见水不是水”作“万水千山总是情”的理解就勉为其难了。

图 1-1 就是一个典型的 MPEG4 场景描述。图 1-2 说明了场景可由对这些自然或合成的二维或三维的原始目标分层次地树型进行描述的方法。

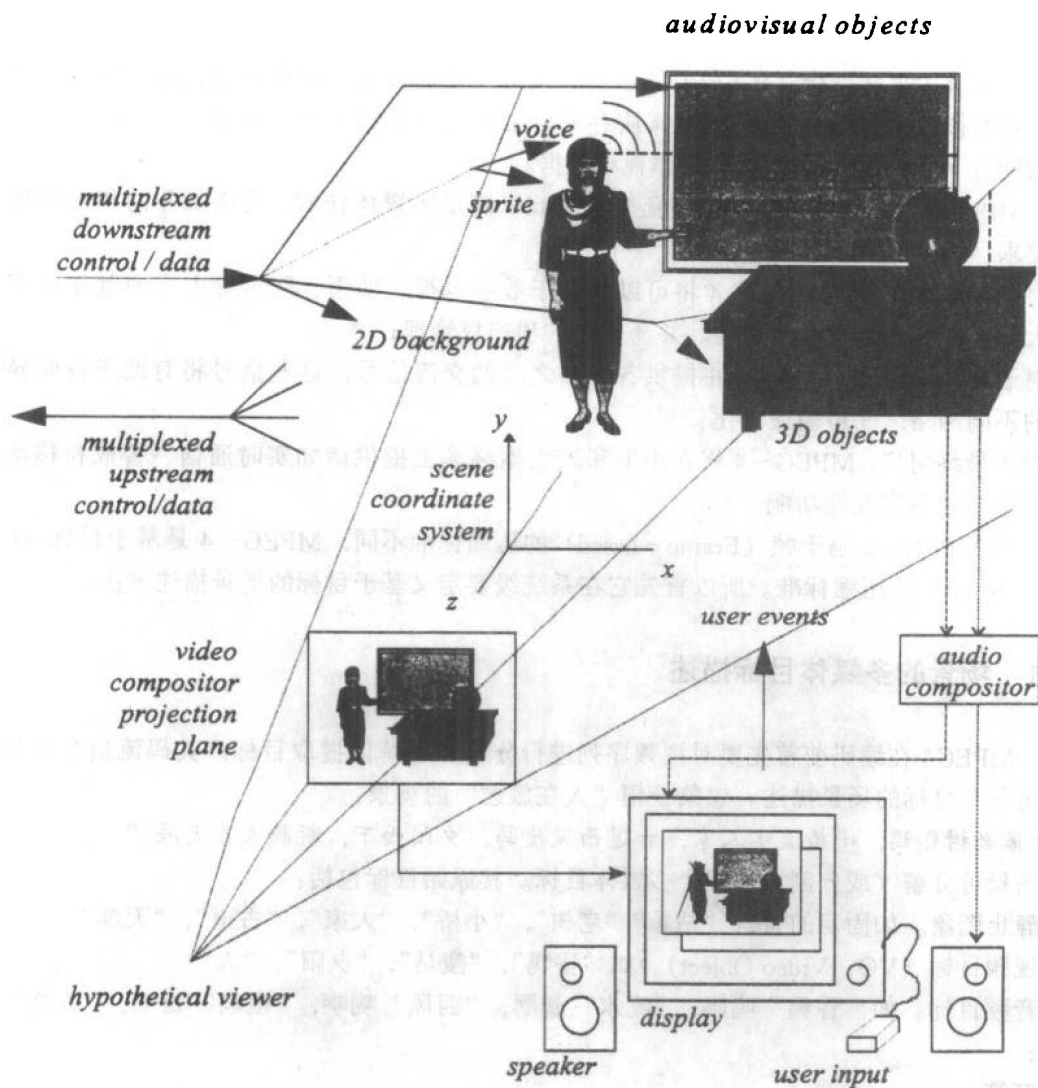


图 1-1 MPEG-4 场景

原始目标通过如图 1-3 所示的 MPEG-4 系统级模型进行发布，并在最终用户端进行合成（如图 1-4）。

相应地，MPEG4 解码器系统级模型如图 1-5 所示。

接下来我们来大致看一下 MPEG-4 视频编码。

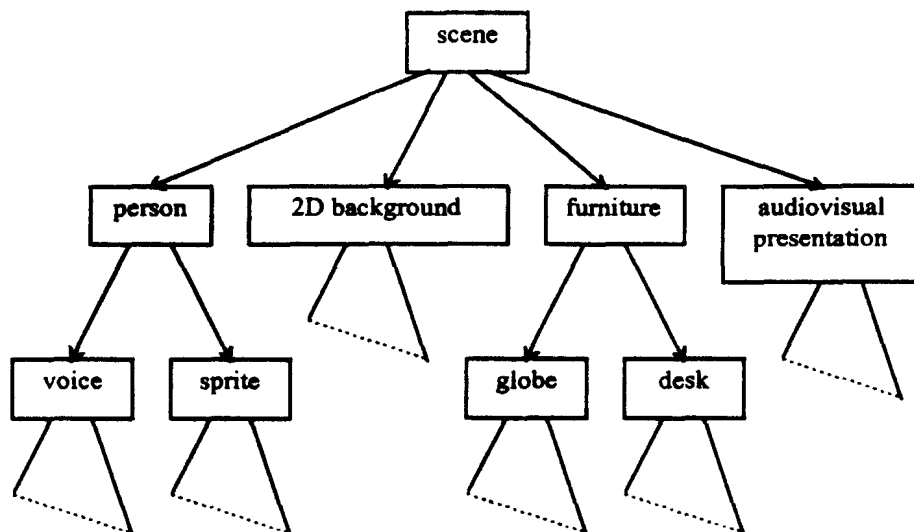


图 1-2 场景的层次化树型描述

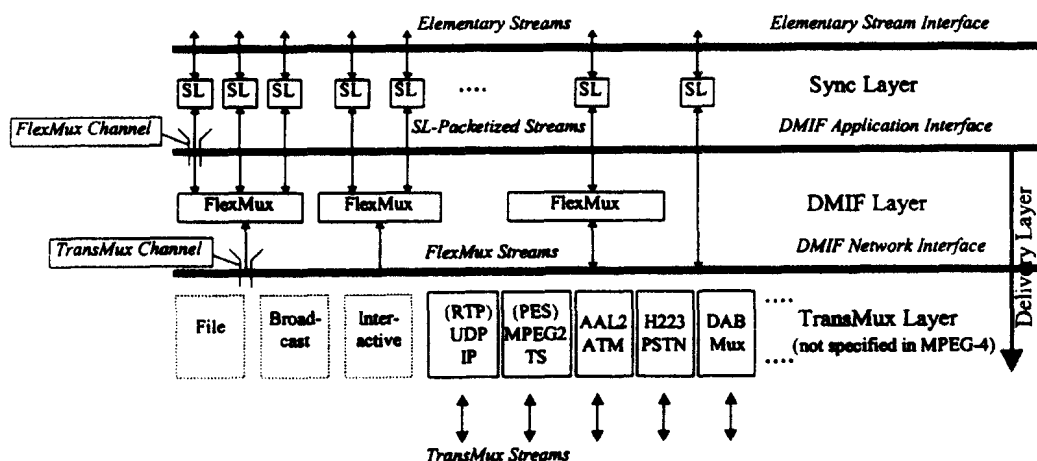


图 1-3 MPEG4 系统级模型

1.2 视频目标编码

MPEG-4 视频目标编码为多媒体诸多应用的视频目标高效存储，传输和操作提供了标准化工具。其特点是：

- 1) 普遍适用性。无论是电视，通信网，计算机都适用；
- 2) 基于内容的交互性。用户可以随机访问和操作目标的内容；
- 3) 高码效率与低码率。码率不超过 64kb/s；

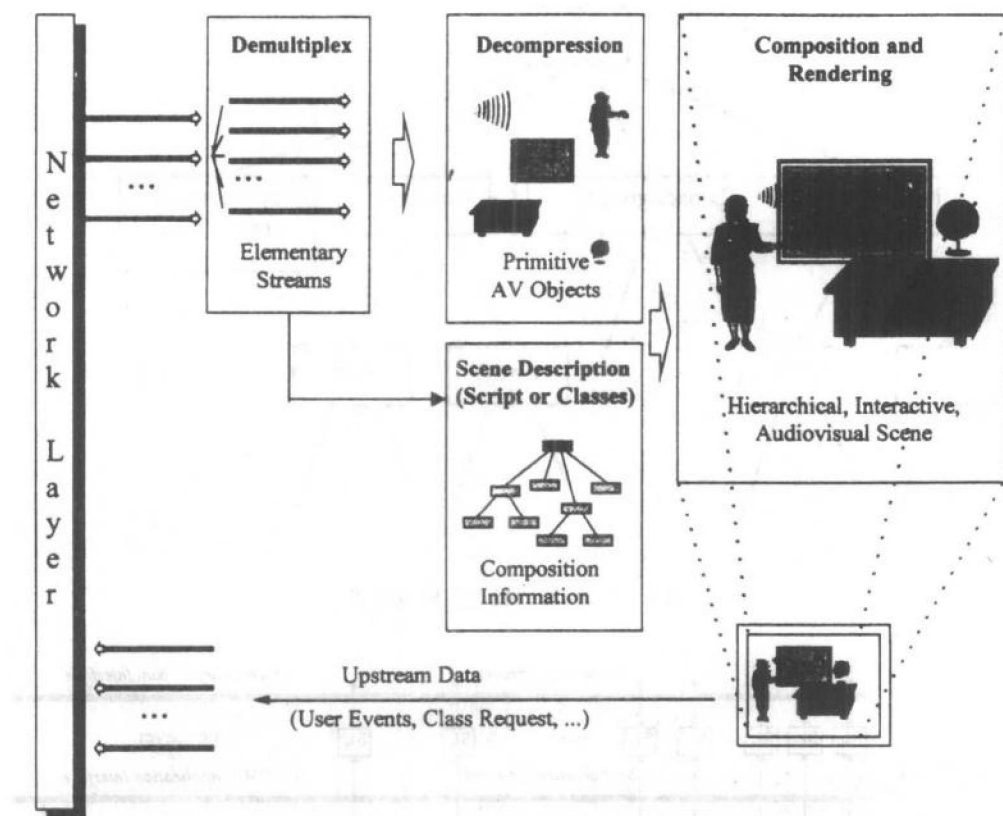


图 1-4 MPEG-4 的场景重建

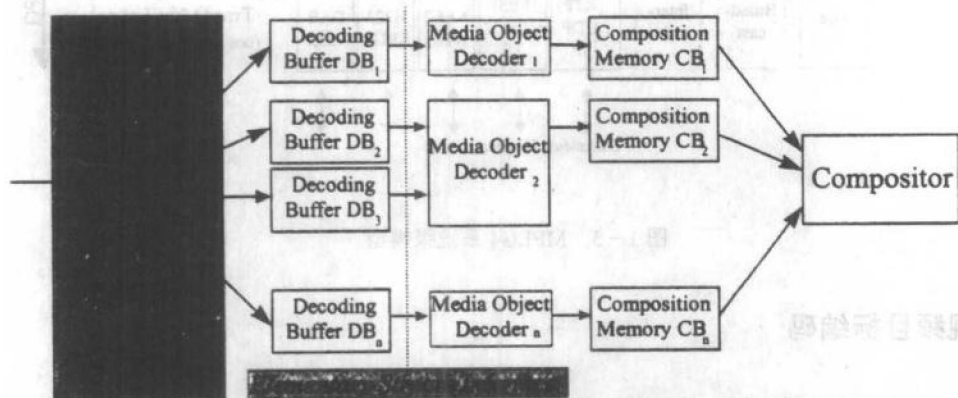


图 1-5 MPEG4 解码器模型

- 4) 高抗错鲁棒性;
- 5) 自然目标和人工合成目标共存;
- 6) 时间和空间的伸展性。

• 4 •