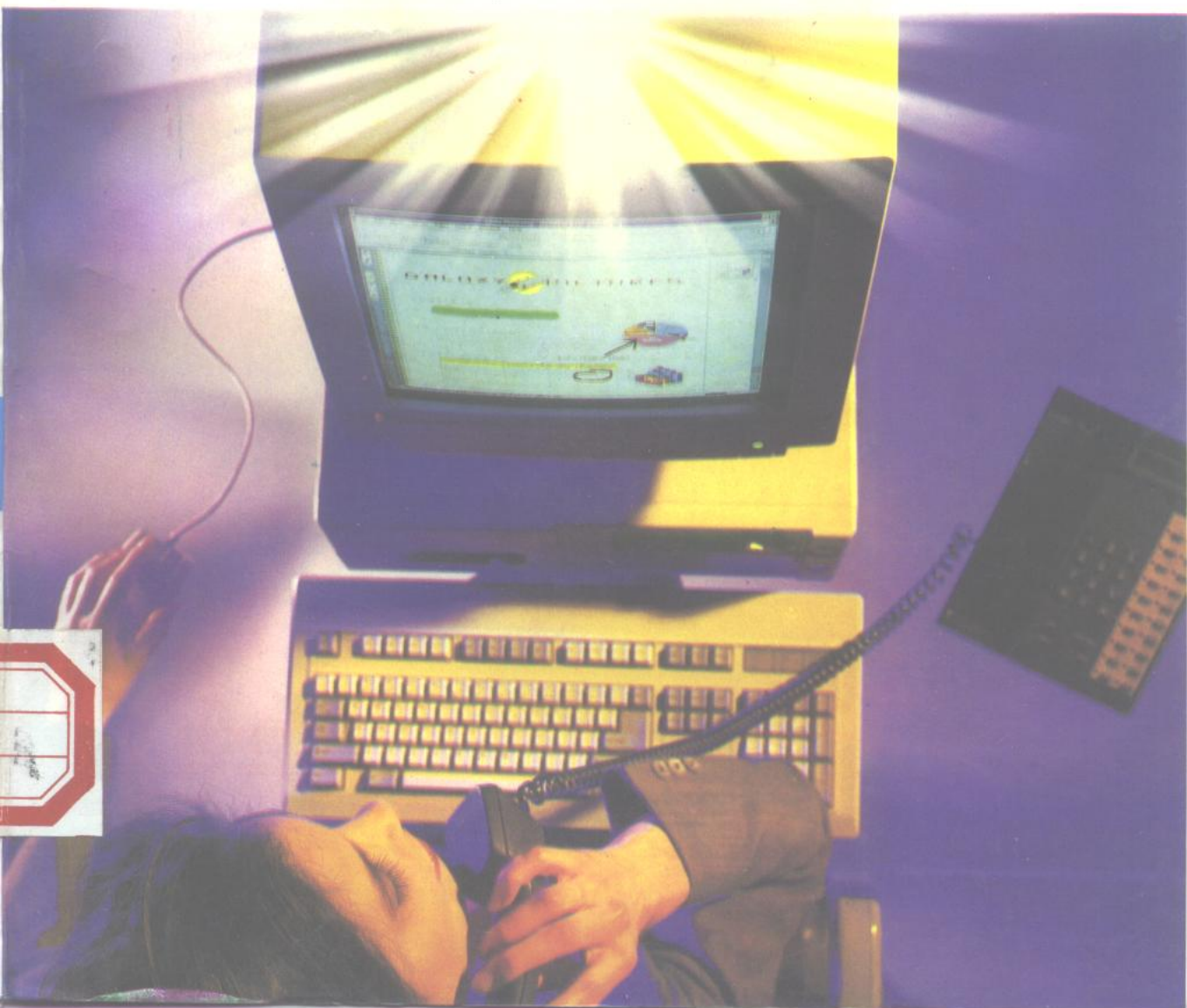


多媒体计算机技术丛书

张福炎主编

运动图像及其伴音 的压缩与编码—— MPEG 1

余崇智 李存珠 张志力 华一满 编译 张福炎 审校



南京大学出版社

(苏)新登字 011 号

内 容 简 介

数字视频图像的压缩与编码技术作为多媒体计算机的一项关键技术,有着非常广泛的应用前景。例如,在电视会议、数字式视频广播、可视电话等的实用化等方面都要应用这项技术。

本书介绍由 CCITT 和 ISO 的运动图像专家组(简称 MPEG)提出的一种运动图像及其伴音的压缩与编码标准——MPEG1,其编号为 ISO/IEC 11172。

本书包含该标准的:第一部分,系统(即 ISO/IEC 11172-1);第二部分,视频(即 ISO/IEC 11172-2);第三部分,音频(即 ISO/IEC 11172-3)。

本书对于我国从事多媒体计算机、数字通信、数字音像技术、电子出版等领域的科研、教学、产品开发的专业技术人员有重要的参考价值。

多媒体计算机技术丛书

运动图像及其伴音的压缩与编码——MPEG1

余崇智 李存珠 张志力 华一满 编译

张福炎 审校

*

南京大学出版社出版

(南京大学内 邮政编码:210093)

江苏省新华书店发行 江苏省地质测绘院印刷厂印刷

*

开本: 787×1092 1/16 印张: 18 字数: 450 千

1995 年 10 月第 1 版 1995 年 10 月第 1 次印刷

印数: 1—6000

ISBN 7-305-02891-6/TP·138

定价: 25.00 元

《多媒体计算机技术丛书》编委会

主 编 张福炎

编 委 蔡士杰 余崇智 唐 棠 潘金贵

宋顺林 潘树陆 李存珠 黄宜华

袁春风 张志力 武港山

《多媒体计算机技术丛书》序

多媒体计算机技术近年来发展异常迅速,它已成为计算机研究、开发及应用领域的一个热点。多媒体技术是计算机、影像处理、语音技术、通讯技术等高度结合的一个产物,它是推动当今计算机技术发展的巨大动力之一。多媒体技术的广泛应用,必将为计算机产业的大发展提供机会。美国苹果电脑公司总裁 John Scully 预言:“多媒体计算机会像 PC 机在 80 年代独领风骚那样,改变 90 年代世界的面貌。”

多媒体技术把文字、图形、图像、动画、音频和视频集成到计算机中,使人们能更加自然、更加有效地使用计算机。人机交互的这种变化极大地拓展了我们的信息空间,满足了人们把多媒体信息做统一处理的需要。

在普通的 386 以上的 PC 机上,装上声音卡、视频卡,配上 CD-ROM 驱动器等所谓的多媒体升级套件,就变成了一台多媒体 PC 机,简称 MPC。如有可能,再挂上 B-ISDN 网(宽带综合业务数字通信网),在多媒体 Windows 的支持下,和音响设备、电视设备联在一起,我们便可以欣赏、编辑、创作立体声音乐、电视图像等,使你进入一个多媒体世界。

我国从事计算机研究、开发及应用的广大人员对多媒体技术极为关注,对它的发展与应用表现出很大热情。但由于这一技术较新,又是多学科交叉,目前有关这方面的技术资料、教材、参考书等还不多见。为帮助国内读者学习和掌握多媒体技术,推动和促进多媒体技术的开发与应用,南京大学多媒体计算机研究所张福炎教授组织编写了这套《多媒体计算机技术丛书》,并由南京大学出版社出版。该丛书包括:

- 1.《多媒体计算机技术概论》;
- 2.《CD 光盘的原理与应用》;
- 3.《数字声音处理及其在多媒体计算机中的应用》;
- 4.《连续色调静止图像的压缩与编码——JPEG》;
- 5.《运动图像及其伴音的压缩与编码——MPEG1》;
- 6.《多媒体计算机软件开发环境与工具》;
- 7.《多媒体软件写作工具及应用》;
- 8.《多媒体计算机人机接口》;
- 9.《计算机动画技术》。

根据这一领域技术发展的需要,我们还将不断地充实、完善本丛书的内容,力求及时反映国内外的最新发展动态和研究成果,以飨读者。

多媒体计算机技术在理论上、技术上以及工程应用方面有许多问题需要研究解决,只要我们深入持久地开展工作,联系应用实际,加强国际合作与交流,就一定能取得可喜的成果,必将为我国的经济建设做出更大的贡献。

《多媒体计算机技术丛书》编委会

1994 年 10 月

前 言

数字视频图像的压缩编码技术有着广泛的应用前景和重大意义,它不仅在多媒体计算机中是一项关键技术,而且也使电视会议、数字式视频广播、可视电话等进入实用阶段。为降低编码/解码器的成本,并使不同厂家生产的设备相互间具有互操作性(interoperability),CCITT 和 ISO 的运动图像专家组(Motion Picture Experts Group,简称 MPEG)提出了一种运动图像及其伴音的压缩编码标准,俗称 MPEG。

MPEG 的第一阶段的成果是 MPEG1,它的编号为 ISO/IEC 11172。它规定了视频信息与伴音信息经压缩之后的数据速率上限为 1.5Mbit/s,从而可以在 CD-ROM、硬盘、可写光盘、数字音频磁带(DAT)等介质上进行存贮,也可以在局域网、ISDN(综合业务数字通信网)网上进行视频与伴音信息的传输。

ISO(国际标准化组织)和 IEC(国际电子技术委员会)为世界范围内的标准化工作组成专门的系统。ISO 或 IEC 的成员国通过各个组织建立的技术委员会,积极参与特定技术领域的国际标准的起草工作。ISO 和 IEC 技术委员会之间在相互感兴趣的领域进行合作。其他一些与 ISO 和 IEC 保持联系的国际组织、政府和非政府机构也参与其中的工作。

在信息技术领域,ISO 和 IEC 建立了联合技术委员会 ISO/IECJTC1。被联合技术委员会采纳的国际标准草案在成员国范围内投票表决。作为国际标准的出版物,至少要得到 75%的成员国赞成票,才能得以通过。

国际标准 ISO/IEC 11172 是由联合技术委员会 ISO/IECJTC1(信息技术)的分委员会 SC29(音频、图像、多媒体和超媒体信息的编码表示)提供的。

ISO/IEC 11172 在总标题“信息技术——上限为 1.5Mbit/s 的数字存贮媒体上的运动图像及伴音的编码”之下,包含下列四个部分:

- 第一部分:系统;
- 第二部分:视频;
- 第三部分:音频;
- 第四部分:一致性测试。

各个部分均有若干附录,有些附录是标准的组成部分,有些附录仅仅是资料性的,供使用时参考。

ISO/IEC 11172 标准对我国广大从事多媒体计算机、数字通信、数字音像技术、电子出版等领域的科研、教学、产品开发的专业技术人员有重要的参考价值。本书对标准的第一、第二和第三部分进行了翻译、整理,特别是第三部分(音频)从便于学习和理解的角度全面进行了改写,但技术内容与标准文本完全一致。在编排上各部分的正文和附录独立编章,为的是尽量贴近原标准文本。

本书第一部分由李存珠翻译,第二部分由张志力翻译,第三部分由余崇智、华一满翻译,全书由张福炎审校、整理、定稿。

限于编者水平,书中错误和不妥之处敬请读者不吝批评指正。

张 福 炎

1994 年 10 月 1 日

目 录

第一部分 系统(ISO/IEC 11172-1)

第一章 引言	3
1.1 复合流操作(包层)	4
1.2 单个流操作(组层)	4
1.2.1 分流	4
1.2.2 同步	5
1.2.3 与压缩层的关系	5
1.3 系统参考解码器	5
第二章 概述	6
2.1 范围	6
2.2 参考资料	6
第三章 技术原理	7
3.1 定义	7
3.2 符号和缩写	13
3.2.1 算术运算符	13
3.2.2 逻辑运算符	14
3.2.3 关系运算符	14
3.2.4 按位运算符	14
3.2.5 赋值	15
3.2.6 助记符	15
3.2.7 常量	15
3.3 比特流语法的描述方法	15
3.4 需求	17
3.4.1 编码结构与参数	17
3.4.2 系统目标解码器	18
3.4.3 系统流语法的规范	21
3.4.4 语法中各字段的语义定义	23
3.4.5 对多路复合流语义的约束	28
3.4.6 约束系统参数流	29
附录 A 系统编码层的描述	30
A.1 概述	30
A.2 编码器操作	30
A.2.1 自由度	30
A.2.2 同步	31
A.2.3 多路复合	32

A. 2. 4	由解码器缓冲所引起的编码器约束	33
A. 2. 5	流特性	34
A. 2. 6	填料流	34
A. 2. 7	专用数据的插入	34
A. 3	解码器操作	35
A. 3. 1	解码器的同步	35
A. 3. 2	解码器的启动同步	37
A. 3. 3	解码器中的缓冲器管理	39
A. 3. 4	时间标识	39
A. 4	用于 CD-ROM 多路复合的参数	39
A. 5	ISO/IEC 11172 多路复合流的例子	40
A. 5. 1	声频	40
A. 5. 2	视频	41
A. 5. 3	多路复合策略	41
A. 5. 4	系统时钟参考(SCR)	42
A. 5. 5	展现时间戳(PTS)	42
A. 5. 6	解码时间戳(DTS)	43
A. 5. 7	缓冲器容量	44
A. 5. 8	遵守系统目标解码器的原则	44
A. 5. 9	样板数据流	46
A. 6	ISO/IEC 11172 多路复合流的结构图	51

附录 B 专利所有者一览表(略)

第二部分 视频(ISO/IEC 11172-2)

第一章	引言	55
1. 1	目的	55
1. 1. 1	编码参数	55
1. 1. 2	受限参数概要	55
1. 2	算法概述	56
1. 2. 1	时域处理	56
1. 2. 2	运动表示——宏块	57
1. 2. 3	减少空域冗余	57
1. 3	编码处理	57
1. 4	解码处理	59
1. 5	编码视频比特流的结构	59
1. 6	算法支持的功能特色	60
1. 6. 1	随机存取	60
1. 6. 2	快速搜索	60
1. 6. 3	反向播放	60
1. 6. 4	差错坚固性	60
1. 6. 5	编辑	60

第二章 概述	61
2.1 范围	61
2.2 参考资料	61
第三章 技术原理	62
3.1 定义	62
3.2 符号和缩写	62
3.3 比特流语法的描述方法	62
3.4 需求	62
3.4.1 编码结构和参数	62
3.4.2 编码的视频比特流语法规范	64
3.4.3 视频比特流语法的语义	69
3.4.4 视频解码过程	77
附录 A 8 * 8 的离散余弦逆变换	86
附录 B 可变长码表	87
B.1 宏块地址	87
B.2 宏块类型	87
B.3 宏块花色	88
B.4 运动矢量	89
B.5 DCT 系数	89
附录 C 视频缓冲检验器	93
C.1 视频缓冲检验器	93
附录 D 视频编码指南	95
D.1 引言	95
D.2 概述	95
D.2.1 视频概念	95
D.2.2 MPEG 视频压缩技术	96
D.2.3 比特流的层次结构	98
D.2.4 解码器概述	99
D.3 预处理	101
D.3.1 CCIR 601 视频到 MPEG SIF 的转换	101
D.3.2 从电影胶片转换	104
D.4 模型解码器	105
D.4.1 模型解码器的目的	105
D.4.2 解码器模型	105
D.4.3 缓冲大小和延迟	106
D.5 MPEG 视频比特流语法	106
D.5.1 序列	106
D.5.2 画面组	111
D.5.3 画面	113
D.5.4 片	116
D.5.5 宏块	119
D.5.6 块	120

D.6 MPEG 视频编码	120
D.6.1 速率控制和自适应量化	120
D.6.2 运动估测和补偿	121
D.6.3 I 画面的编码	128
D.6.4 P 画面的编码	135
D.6.5 B 画面的编码	141
D.6.6 D 画面的编码	144
D.6.7 画面速率较低时视频的编码	144
D.7 MPEG 视频的解码	145
D.7.1 正向播放时画面的解码	145
D.7.2 快速播放时画面的解码	146
D.7.3 暂停和步进方式的解码	146
D.7.4 反向播放时的画面解码	146
D.8 后处理	147
D.8.1 编辑	147
D.8.2 再采样	147
附录 E 参考文献	150
附录 F 专利所有者一览表(略)	

第三部分 声频(ISO/IEC 11172-3)

第一章 综述	153
1.1 范围	153
1.2 系统概述	153
1.2.1 编码器	153
1.2.2 层	154
1.2.3 存贮	154
1.2.4 解码	154
第二章 声频编码过程	155
2.1 概述	155
2.1.1 滤波器组	155
2.1.2 心理声学模型	155
2.1.3 比特或噪声分配	156
2.1.4 比特流构成	156
2.1.5 输入高通滤波器	156
2.2 分析子带滤波器	156
2.3 层 I 编码	161
2.3.1 比例因子	162
2.3.2 比特分配	162
2.3.3 子带样点	164
2.3.4 辅助数据	164
2.3.5 格式化	164

2.4 层Ⅰ编码	165
2.4.1 比例因子	165
2.4.2 比特分配	166
2.4.3 子带样点	172
2.4.4 辅助数据	173
2.4.5 格式化	173
2.5 层Ⅱ编码	173
2.5.1 心理声学模型	173
2.5.2 混合滤波器组的分析部分	187
2.5.3 平均可用比特数的计算	188
2.5.4 频域样点的量化和编码	189
2.5.5 辅助数据	189
2.5.6 格式化	189
2.6 层Ⅲ迭代循环	191
2.6.1 引言	191
2.6.2 预备步骤	192
2.6.3 外迭代循环(畸变控制循环)	194
2.6.4 内迭代循环(比率控制循环)	196
2.7 联合立体声编码	209
2.7.1 层Ⅰ、层Ⅱ强度立体声编码	209
2.7.2 层ⅢMS立体声和强度立体声编码	211
2.8 差错的比特灵敏度和差错隐蔽	213
2.8.1 概述	213
2.8.2 层Ⅰ和层Ⅱ	213
2.8.3 层Ⅲ	214
2.8.4 差错隐蔽	214
第三章 音频解码过程	216
3.1 概述	216
3.2 层Ⅰ	217
3.2.1 子带样点的再量化	218
3.2.2 综合子带滤波器	218
3.3 层Ⅱ	218
3.3.1 比特分配解码	218
3.3.2 比例因子选择信息解码	218
3.3.3 比例因子解码	219
3.3.4 子带样点再量化	219
3.3.5 综合子带滤波器	219
3.4 层Ⅲ	223
3.4.1 解码	224
3.4.2 侧信息	227
3.4.3 主数据开始	228
3.4.4 缓冲器考虑	228
3.4.5 比例因子	228

3.4.6 霍夫曼解码	229
3.4.7 再量化器	229
3.4.8 重新排序	230
3.4.9 立体声处理	230
3.4.10 综合滤波器组	231
第四章 音频比特流语法	235
4.1 音频序列	235
4.1.1 层 I 和层 II	235
4.1.2 层 III	235
4.2 音频帧	235
4.3 标题	236
4.4 差错检验	239
4.5 音频数据层 I	240
4.6 音频数据层 II	241
4.7 音频数据层 III	243
4.7.1 音频数据	243
4.7.2 主数据	247
4.7.3 霍夫曼码	249
4.8 辅助数据	250
第五章 心理声学模型	251
5.1 心理声学模型 1	251
5.1.1 概述	251
5.1.2 FFT 分析	251
5.1.3 声压级测定	252
5.1.4 安静状态阈值	252
5.1.5 判决有调和无调成分	252
5.1.6 有调和无调掩蔽成分的抽取	254
5.1.7 单独掩蔽阈的计算	254
5.1.8 总掩蔽阈 LT_0 计算	255
5.1.9 决定最小掩蔽阈值	255
5.1.10 计算信号-掩蔽-比率	255
5.2 心理声学模型 2	265
5.2.1 概述	265
5.2.2 符号注释	265
5.2.3 扩展函数	265
5.2.4 阈值计算	266

第一部分 系 统

(ISO/IEC 11172-1)

第一章 引言

本系统规范论述了这样一个问题,即如何将符合本标准视频和音频部分的一条或多条数据流与定时信息相结合,形成单一的复合流。组合成一个流之后,这些数据就适合于数字存贮或数字传输了。本规范所使用的语法和语义规则使得在许多情况下接收或检索到的流能够同步播放而不会出现解码缓冲器的上溢或下溢。本规范的语法和语义规则的作用范围并不相同:语法规则仅适用于系统层编码,并不扩大到视频和音频规范的压缩层编码;相反,语义规则则对整个复合流都起作用。

本系统规范并没有规定编码器或解码器的体系结构或实现方法。但是,比特流的许多性质对编码器和解码器必须满足的功能和性能提出了要求。例如,编码器必须满足最小的时钟容差要求。尽管如此,在设计和实现编码器和解码器时仍有许多自由度。

图 1.1 描述了一个音频/视频解码系统原型,它图示了 ISO/IEC 11172 解码器的功能。该体系结构并不是唯一的——包括解码器定时控制在内的许多系统解码器功能可以完全等价地分散到基本流解码器和特定的介质解码器中去——但是该图有助于讨论。这个解码器原型的设计并不意味着在设计一个 ISO/IEC 11172 解码器时有任何强制性的要求。实际上,非音频/视频数据也是允许的,只是在图中没有表示出来而已。

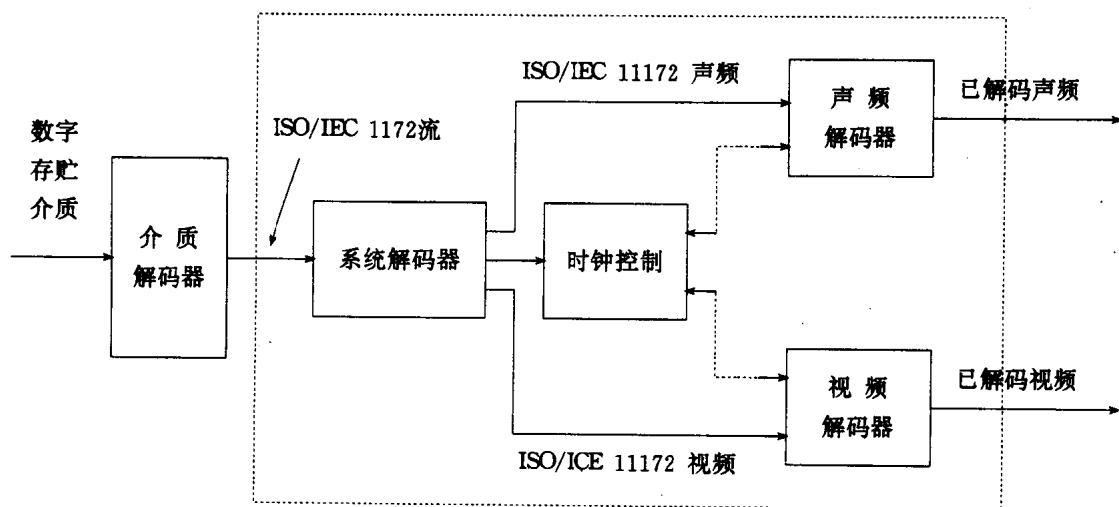


图 1.1 ISO/IEC 11172 解码器原型

图 1.1 中所示的 ISO/IEC 11172 解码器原型由系统解码器,视频解码器和音频解码器所组成。这些解码器分别符合于 ISO/IEC 11172 的第 1 部分、第 2 部分和第 3 部分的规范。解码器中,一条或多条音频和视频流多路复合而成的代码假设以介质特定的格式存贮在数字存贮介质(DSM,digital storage medium)上或存贮在网络上。本国际标准既不规定介质特定的格式,介质解码器也不是 ISO/IEC 11172 解码器原型中的组成部分。

解码器原型接受 ISO/IEC 11172 多路复合流作为输入,借助于系统解码器从多路复合流

中抽取定时信息。系统解码器对该输入流进行分流处理,所产生的基本流作为视频解码器和音频解码器的输入,视频解码器和音频解码器的输出分别是已解码的视频信号和音频信号。系统解码器、视频解码器、音频解码器以及介质解码器之间的定时信息流虽然在图中没有表示出来,但应包括在设计之中。视频解码器和音频解码器相互之间,以及与数据存贮介质之间是使用定时信息进行同步的。

ISO/IEC 11172 多路复合流构造成为两层:系统层和压缩层。系统解码器的输入流是压缩层外围的系统层,输入到视频解码器和音频解码器的输入流则为压缩层。

系统解码器执行两类操作,一类是作用在整个 ISO/IEC 11172 多路复合流上的操作,称为“复合流操作”(multiplex-wide operations);另一类是作用在单个基本流上的操作,称为“特定流操作”(stream-special operations)。ISO/IEC 11172 系统层分成两个子层,一个子层称为“包”(pack),是多路操作的作用对象,另一个子层称为“组”(packet),它是用于特定流操作的。

1.1 复合流操作(包层)

多路复合流操作包括 DSM 读出数据的协调,时钟的调节,以及缓冲器的管理等任务。这些任务都是密切相关的。如果从 DSM 读出数据的速率是可控制的,那么就可以调节 DSM 的数据速率,使得解码器中的缓冲器既不会上溢也不会下溢。但是,如果 DSM 的传输速率不是可控制的,那么,基本流解码器就必须使其时钟服从于 DSM 的控制以避免上溢和下溢。

ISO/IEC 11172 多路复合流由包组成,这些包的标题信息有助于上述任务的完成。在包的标题信息中给出了从 DSM 送来的每个字节进入系统解码器的预期时间,这个到达时刻表对时钟校正和缓冲器管理也能起参考作用。虽然解码器不必精确地按照时刻表的要求进行操作,但若有偏离则必须进行补偿。

另一个复合流操作是解码器应能确定解码所需资源。每个 ISO/IEC 11172 多路复合流的第一个包中都包含了一些参数来协助解码器完成此任务。例如,复合流的最大数据速率以及视频通道的最大数目等等。

1.2 单个流操作(组层)

特定流操作主要包括:

- (1)分流;
- (2)多条基本流的同步播放。

1.2.1 分流

在编码的时候,是由多个基本流通过复合操作形成 ISO/IEC 11172 多路复合流。除了 ISO/IEC 11172 音频流和视频流以外,基本流可以包括专用流、备用流以及填料流。这些流都被临时性的划分成许多组,组与组之间保持原顺序。一个组中只能包含来自同一个基本流中的编码字节。

组的长度可以是固定的,也可以是可变的,但都必须遵循 3.4.3 节、3.4.5 节和 3.4.6 节中的约束。

解码时,需要对 ISO/IEC 11172 多路复合流进行分流处理,以重构基本流,这可以借助于

组标题中的流-标识符来完成。

1.2.2 同步

多个基本流之间的同步是通过使用 ISO/IEC 11172 多路复合流中的展现时间戳得以实现的。该时间戳以 90kHz 为单位。同步播放 N 个流的方法是使所有播放的流都使用同一个时间基准进行调准,而不是相互之间互相适配。这一个时间基准可以是 N 个解码器时钟之中的某一个,也可以是 DSM 或通道时钟,或者是某个外部时钟。

由于展现时间戳作用于各个基本流的解码操作,所以这些时间戳就应该存放于组这一层中。如果编码器在数据获取时就记录下时间值(称为“时间戳”)并把时间戳与相关的编码数据一起传输到解码器,解码器再利用这些时间戳来安排信息展现的时间,那么就能实现端到端的同步,即解码端输出信息的同步情况与编码端的输入情况完全一致。

在多路复合数据流中,采用 DSM 中的时间戳也可以达到同步的目的。

1.2.3 与压缩层的关系

在某种意义上来说组层是独立于压缩层的,但也并非完全如此。组不要求以压缩层中的开始码作为其开始位置,所以它们是相互独立的。例如,一个视频组可以开始于该视频流的任何字节处。然而,在组标题中的时间戳是压缩层中一些结构(即展现单元)的展现时间,在这个意义上组与压缩层又是相关的。

1.3 系统参考解码器

ISO/IEC 11172 的第 1 部分利用“系统目标解码器”(system target decoder, STD)给定时和缓冲的关系提供了一种形式化的表示方法。由于 STD 是以 ISO/IEC 11172 中所定义的一些字段(如缓冲器容量等)作为其参数进行描述的,所以每个 ISO/IEC 11172 多路复合流都有它自己的一组 STD 参数。编码器应确保所产生的比特流能按正常的速度在相应的 STD 解码器上进行正向播放。物理解码器可以假设比特流在 STD 上工作是正确的,但是该物理解码器必须对它与 STD 之间的差异作出补偿。

第二章 概 述

2.1 范 围

ISO/IEC 11172 的这一部分对系统层编码给出了规定。其主要目的是为了支持把 ISO/IEC 11172-2 中定义的视频编码方法和在 ISO/IEC 11172-3 中定义的音频编码方法组合起来。系统层支持五个基本的功能:

- a. 多条压缩流同步演播;
- b. 多条压缩流交错组合成为一个单一流;
- c. 为启动演播而对缓冲区进行初始化;
- d. 连续的缓冲器管理;
- e. 时间标识。

一个 ISO/IEC 11172 多路复合比特流由两层构成:最外面的层是系统层,最内层是压缩层。系统层提供了为在一个系统中使用一条或多条压缩数据流所必须的功能。本规范的视频和音频部分分别定义了视频和音频数据的压缩编码层。本规范没有定义其他类型数据编码,但它们若能遵循 3.4 节所定义的一些约束,则系统也可予以支持。

2.2 参 考 资 料

下面的国际标准包含有在正文中所引用的有关内容,它们也是 ISO/IEC 11172 这个部分的组成部分。

- ISO/IEC11172-1 本标准的第一部分,系统。
- ISO/IEC 11172-2 本标准的第二部分,视频。
- ISO/IEC 11172-3 本标准的第三部分,音频。
- CCIR601-2 建议 数字电视演播的编码参数。
- CCIR624-4 报告 黑白和彩色电视系统的特性。
- CCIR648 建议 音频信号的记录。
- CCIR955-2 报告 用于便携式和移动式接收装置的卫星声音广播。
- CCITTJ. 17 建议 使用在声音节目线路上的 pre-emphasis 技术。
- IEEE P1180/D2 1990 实现 8×8 离散余弦逆变换的规范(标准草案)。
- IEC 出版物 908 1987 CD 数字声响系统。