



中华人民共和国国家标准

GB/T 20090.1—2012

信息技术 先进音视频编码 第 1 部分：系统

Information technology—Advanced coding of audio and video—
Part 1: System

2012-12-31 发布

2013-06-01 实施

中华人民共和国国家质量监督检验检疫总局 发布
中国国家标准化管理委员会

目 次

| | |
|--|-----|
| 前言 | III |
| 引言 | IV |
| 1 范围 | 1 |
| 2 规范性引用文件 | 1 |
| 3 术语和定义 | 2 |
| 4 符号和缩略语 | 7 |
| 5 位流语法的描述方式 | 9 |
| 6 传输流位流要求 | 10 |
| 7 节目流位流的要求 | 56 |
| 8 节目和节目元素描述子 | 70 |
| 9 对多路复用流语义的约束 | 85 |
| 10 与 GB/T 17975 的兼容性 | 88 |
| 附录 A (资料性附录) 系统时间信息时钟恢复模型及应用指南 | 89 |
| 附录 B (资料性附录) 系统符合性 | 98 |
| 附录 C (资料性附录) 引入抖动的网络与 AVS 解码器的接口 | 99 |
| 附录 D (资料性附录) 语法的图形表示 | 102 |
| 附录 E (资料性附录) 通用信息 | 106 |
| 附录 F (资料性附录) 拼接传输流 | 107 |
| 附录 G (资料性附录) 数据传输应用 | 110 |
| 附录 H (资料性附录) 数字存储媒体命令与控制(DSM-CC) | 111 |
| 附录 I (规范性附录) CRC 解码器模型 | 121 |
| 附录 J (资料性附录) 系统解复用过程 | 122 |
| 附录 K (资料性附录) 系统复用过程 | 125 |
| 附录 L (资料性附录) 节目特定信息 | 129 |
| 附录 M (资料性附录) 私有数据 | 138 |

前 言

GB/T 20090《信息技术 先进音视频编码》分为以下 9 个部分：

- 第 1 部分：系统；
- 第 2 部分：视频；
- 第 3 部分：音频；
- 第 4 部分：符合性测试；
- 第 5 部分：参考软件；
- 第 6 部分：数字版权管理；
- 第 7 部分：移动视频；
- 第 8 部分：在 IP 网络上传输 AVS；
- 第 9 部分：AVS 文件格式。

本部分为 GB/T 20090 的第 1 部分。

本部分按照 GB/T 1.1—2009 给出的规则起草。

本部分由中华人民共和国工业和信息化部提出。

本部分由全国信息技术标准化技术委员会(SAC/TC 28)归口。

本部分起草单位：中国科学院计算技术研究所、清华大学、北京算通科技发展有限公司、中国电子技术标准化研究院、北京大学、华中科技大学、北京邮电大学、联合信源数字音视频技术(北京)有限公司、展讯通信(上海)有限公司。

本部分主要起草人：陈熙霖、杨士强、高鹏飞、霍龙社、高麟鹏、丁亚强、杨志杰、林福辉、牟伦田、王啸、朱光喜、林永兵、贺玉文、喻莉、王雷、冯博、张桢睿、杨昭辉、杨洁、汪恒晶、高文、黄铁军、卢汉清。

引 言

0.1 综述

GB/T 20090 的本部分论述了如何将一路或多路音频、视频位流或其他基本数据流合成单路或多路复用流,用于存储和传送。系统编码遵循本部分指定的语法和语义规则,并提供了使解码器缓冲区能在一定范围的补偿和接收条件下进行同步解码的信息。

系统编码包含两种形式:传输流和节目流,分别针对不同的应用加以优化。本部分中定义的传输流和节目流提供了编码语法,该语法对于同步解码及展现音频、视频信息是充分且必要的,同时保证了解码器中数据缓冲区不发生上溢和下溢。该语法中应至少包含以下两种信息之一:

- a) 有关展现编码的音频和视频数据的相对时间戳,以及有关数据流本身传送的比特率信息;
- b) 有关解码和展现编码的音频和视频数据的时间戳,以及有关数据流本身传送的时间戳信息。

传输流和节目流都是面向分组包的多路复用流。

单一音频和视频基本流的多路复用过程见图 1。视频和音频数据可按包括 GB/T 20090.2 和 GB/T 20090.3 等部分进行编码,之后,压缩数据被打包以形成 PES 分组包。在形成 PES 分组包的过程中,当独立使用传输流或节目流的 PES 分组包时可能会加入所需的描述信息;而 PES 分组包需要进一步与系统层信息结合形成传输流或节目流时,不包含这一信息。本部分覆盖了图 1 中竖直虚线右边所示的处理过程。

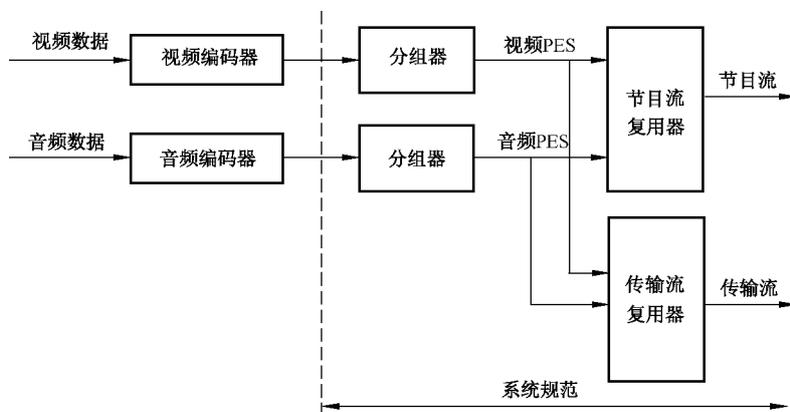


图 1 多路复用过程简图

本部分定义的节目流是由具有共同时间基准的一个或多个 PES 分组包合并而成的单一流。有些应用中要求包含单个节目的基本流是未多路复用的分离流。对这些应用,基本流也可作为分离的节目流编码。每一基本流含一个节目流且具有共同时间基准。在这种情况下,不同流中 SCR 字段的编码值应一致。

所有的基本流和单一节目流都可被同步解码。

节目流被设计用于相对无差错环境中,且适用于诸如交互式多媒体等应用。节目流分组包长度可变。

传输流是由一个或多个节目组成的单一流,且这些节目可包含一个或多个不同时钟基准。由同一个节目中的多个基本流所组成的多个 PES 分组包共享一个时钟基准。传输流是为用于可能出现差错

的有损或有噪媒体中的存储或传输环境而设计的。传输流分组包长度为 188 字节。

节目流和传输流是为不同应用设计的,它们的定义并不严格遵守分层模型,彼此之间可相互转换,但并不互为子集或超集。可以从一个传输流中抽取一个节目的内容并创建一个有效的节目流。该转换可利用 PES 分组包的公共互换格式完成,但并非节目流需要的所有字段都包含在传输流中,有一些需要导出。而在分层模型中,传输流可能横跨多个层,并易于在宽带应用中实现。

本部分描述的语法规则和语义规则的范围是不同的:语法规则仅用于系统层编码,并不延伸到 GB/T 20090.2 和 GB/T 20090.3 中所描述的压缩层编码,而语义规则适用于复用流。

本部分未规定编码器或解码器的体系结构或实现方法,也未对多路复用器或解复用器作相应的规定。然而,位流的性质对编码器、解码器、多路复用器和解复用器提出了功能和性能上的要求,如编码器应满足最小的时钟容差要求,但这些要求并不妨碍编码器、解码器、多路复用器和解复用器的设计与实现的灵活性。

0.2 传输流

传输流是针对在那些可能会出现显著错误(往往表现为位差错或分组包丢失)的环境中进行节目传送和存储而定义的一种流。这些节目包含按照 GB/T 20090.2 和 GB/T 20090.3 等部分定义的编码数据以及其他数据。

传输流的速率可以是恒定或可变的。在任何情况下,所包含的基本流的速率也可以是恒定或可变的。在每一种情况下,流的语法或语义限制是相同的,传输流速率由传输率(transport_rate)字段的值定义,这些传输率字段嵌入在每个流中。

一个包含多个具有独立时间基准的节目传输流会造成总体比特率可变,这会构造和传送这样的位流带来一定困难,见 6.2.2。

只要能够生成一个有效的流,包含一个或多个节目的传输流可以从基本编码数据流、节目流或其他可能包含一个或多个节目的传输流构造生成。

传输流是按照在最小开销的情况下能对传输流执行某些操作的原则而设计的。这些操作包括:

- a) 从传输流的一个节目中获得编码数据、解码并展现,如图 2 所示;
- b) 从传输流的一个节目中抽取传输流分组包并生成一个仅包含该节目的新的传输流作为输出,如图 3 所示;
- c) 从一个或多个传输流中抽取一个或多个节目的传输流分组包并生成一个新的传输流;
- d) 从传输流中抽取一个节目内容并生成包含该节目的一个节目流,如图 4 所示;
- e) 把一个节目流转化为传输流,并在有损环境中传输。然后再重建一个有效的、在某些情况下完全相同的节目流。

图 2 和图 3 描述了以一个传输流作为输入的解复用及解码系统原型。图 2 说明了第一种情况,即一个传输流被直接解复用和解码。传输流构造分为两层:系统层和压缩层。

传输流解码器的输入流由一个包含压缩层的系统层组成。音频解码器及视频解码器的输入流只包含压缩层。

接收传输流的解码器原型的操作既适用于整个传输流(复用流操作),也适用于单个基本流(特定流操作)。传输流系统层被分为两个子层,一个用于复用流操作(传输流分组包),另一个用于特定流操作(PES 分组包层)。

图 2 也给出了一个包括视频和音频的传输流解码器原型以说明解码器的功能,其结构并不是唯一的,但该图有助于理解。解码器时间控制等系统解码器功能可能被等价地实现在基本流解码器或信道特定解码器中。类似地,信道专用解码器测出的错误也可以用多种途径通知独立的音频和视频解码器。这些通信途径并未显示在图中,该解码器原型的设计并不意味着对传输流解码器的设计作出任何必达

的要求。实际上,非音视频数据也是允许的,但并未在图 2 中画出。

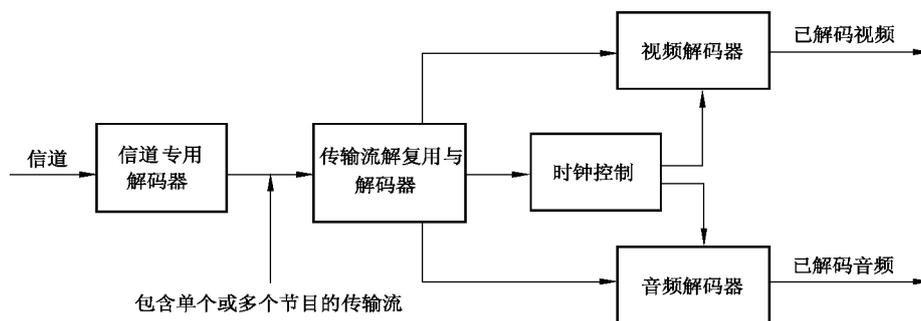


图 2 传输流解复用和解码原型示例

图 3 说明了另一种情况,即一个包含多个节目的传输流被转变为一个只含单个节目的传输流,这种情况下的再复用操作可能需要纠正 transport_rate 值以补偿位流中传输率的变化。

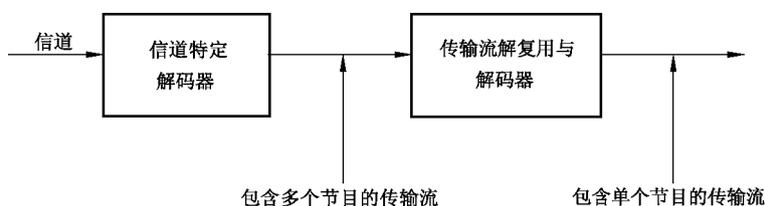


图 3 传输多路复用原型示例

图 4 说明了一个多节目传输流先被解复用再转变为节目流的情况。

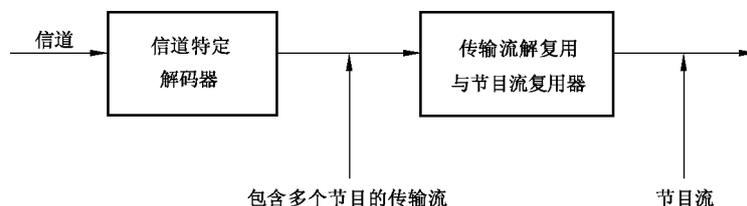


图 4 传输流到节目流转换原型示例

图 3 和图 4 指出,不同类型、不同构造的传输流之间的转换是可能且合理的。在传输流和节目流的语法中都定义了一些特定字段以方便上述转换过程,但并不要求解复用器或解码器的具体实现要包含以上所有功能。

0.3 节目流

节目流是针对在那些出错率很低,且系统编码的处理过程作为主要考虑因素的环境中进行一个节目的传送和存储而定义的一种流,该节目流包含编码数据和其他数据。

节目流的速率可以是恒定的或可变的。在任何情况下,所包含的基本流的速率也是恒定或可变的。在每一种情况下,节目流的语法或语义限制是相同的。节目流速率是由系统时钟参考(SCR)字段与 mux_rate 字段的值和位置所决定的。

图 5 描述了一个音视频节目流解码系统原型。该图用于帮助理解,其结构并不是唯一的——包括解码器时间控制在内的系统解码器功能可能被等价地实现在基本流解码器或信道特定解码器中。该解码器原型的设计并不意味着对节目流解码器的设计作出任何必达的要求。实际上,非音视频数据也是

允许的。

图 5 所示的节目流解码器原型是由系统、视频和音频解码器三部分组成,它们分别符合 GB/T 20090.1、GB/T 20090.2 和 GB/T 20090.3。在该解码器中,单路或多路音视频位流的复用编码表示假定以某种特定信道格式在特定信道中存储或传输。特定信道格式本部分不作规定,特定信道解码也不是本解码器原型的一部分。

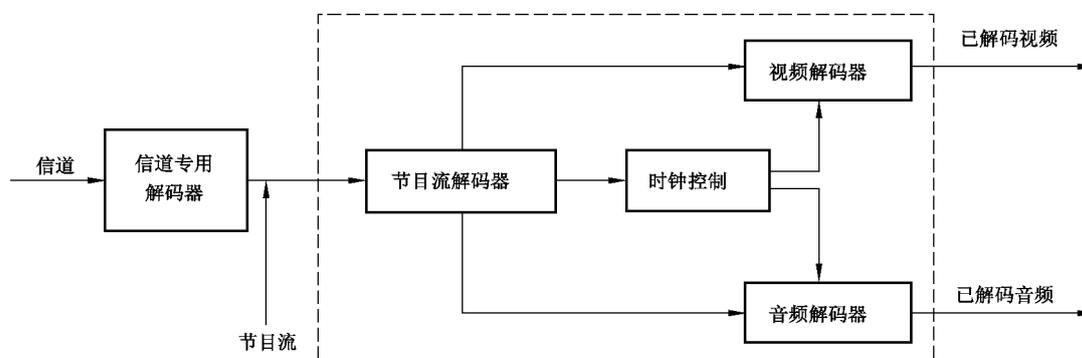


图 5 节目流解码器原型

上述原型解码器接受节目流作为输入,并依靠节目流解码器从流中提取信息。节目流解码器分离复用流,并将由此产生的基本流作为音频和视频解码器的输入。音频和视频解码器的输出是已解码的音频和视频信号。节目流解码器、音频和视频解码器以及信道专用解码器之间的定时信息流应包含在设计中,但并未在图 5 中画出。利用定时信息,音频和视频解码器相互之间及与信道之间可以实现同步。

节目流构造分为两层:系统层和压缩层。节目流解码器的输入流由一个包含压缩层的系统层组成。音频解码器及视频解码器的输入流只包含压缩层。

解码器原型的操作既适用于整个节目流(复用流操作),也适用于单个基本流(特定流操作)。节目流系统层被分为两个子层,一个用于复用流操作(节目流分组包层),另一个用于特定流操作(PES 分组包层)。

0.4 传输流与节目流的转换

利用 PES 分组包,传输流与节目流之间的转换是可能且合理的,这是由包含在本部分的 6.1 和 7.1 中的传输流和节目流规范得出的。在某些限制下,PES 分组包可能直接从一个复用位流的有效数据对应到另一个复用位流的有效数据。如果在所有 PES 分组包中都有 program_packet_sequence_counter,就可能标识出 PES 分组包的正确次序以帮助实现这一功能。

在这两种流中,有关转换所需的其他信息包括基本流之间的关系以表和包头的形式出现。在任何流中这些数据如果出现,转换前后都应是正确的。

0.5 PES 流

正如 6.3.6 中的语法定义所指出,传输流和节目流是从 PES 分组包中逻辑地建立的。PES 分组包被用于传输流与节目流之间的转换。在有些场合进行这种转换时,无需变动 PES 分组包。PES 分组包的长度可能比传输流分组包的长度大得多。

包含相同 stream_id 和基本流的一系列连续的 PES 分组包构成一个 PES 流。当 PES 流分组包用于形成 PES 流时,应在 6.3.8 中所定义的限制下,带有基本流时钟参考(ESCR)字段和基本流速率(ES_

Rate) 字段。PES 流数据应是来自基本流且保持原次序的连续字节。PES 流中不包含某些出现在节目流和传输流中的系统信息,包括在包头、系统头、节目流映射、节目流目录、节目映射表中出现的信息以及在传输流分组包语法中定义的元素。

本部分的 PES 流未定义用于相互交换和交互操作。单个基本流也可打包成传输流或节目流,但应包含必要的系统信息。具有单个基本流的多个传输流或节目流可被构造具有相同的时间基准以传送一个完整的音视频节目。

0.6 定时模型

系统、视频和音频都有一个定时模型,其中从编码器的信号输入到解码器的信号输出之间的端到端延迟是恒定的,该延迟是编码、编码器缓冲、多路复用、传送或存储、解复用、解码器缓冲和展现过程中所有延迟的总和。作为该定时模型的一部分,所有视频图像和音频采样仅展现一次(除非经过特殊编码),且解码器中的视频帧率和音频采样速率与在编码器中保持一致。系统流编码包括了定时信息,以用于实现端到端延迟恒定的系统。实现不严格遵守该模型的解码器也是可能的。但此时的解码器需要负责以一种可接受方式完成以上要求。

所有定时是由一个公共时钟来定义的,这个公共时钟被称为系统时钟。在节目流中,该时钟与视频或音频采样时钟之间可以有确定的比值,也可以有一个与比值略有偏差的工作频率,但仍提供精确的端到端定时和时钟补偿。

在传输流中,为了简化解码器中的采样速率恢复,系统时钟被限制为在任何时刻均与音频和视频采样时钟保持确定比值。

0.7 条件接收

本部分未规定条件接收机制,但 AVS 位流支持对节目流和传输流中的节目进行条件接收。

0.8 复用流操作

复用流操作包括协调信道数据的获取、时钟的调整以及缓冲区的管理。这些操作是紧密相连的。当信道数据传输速率是可控时,可通过调节数据传输速率使解码缓冲区不发生上溢或下溢。否则,基本流解码器的定时应服从信道中接收的数据,以避免上溢或下溢。

节目流由节目流分组包组成,其包头有助于以上任务的完成。包头指定了从信道中送来的每一字节进入节目流解码器的预定时间,这个预定到达时刻表作为时钟校正和缓冲区管理的参考。当解码器不能严格遵守该时刻表时,应对有关偏差作出相应补偿。

类似地,传输流由传输流分组包构成。分组包头指定从信道中送来的每一字节进入传输流解码器的预定时间。该时刻表提供了与上述节目流中完全相同的功能。

另一个复用流操作是确定解码器解码传输流或节目流时所需的资源。每个节目流的第一个包均包含一些参数,来协助解码器完成此功能,其中包括流的最大数据速率以及同步视频信道的最大数目。传输流也包含类似的全局适用的信息。

每个传输流和节目流都包含一些信息,以标识组成一个节目的各基本流的相关特征以及基本流之间的相互关系。这些信息可包括声道中所使用的语言,以及在实现多层视频编码时各层之间的关系。

0.9 单个流操作(PES 分组包层)

基本的 PES 流的操作为:

- a) 解复用;
- b) 多个基本流的同步回放。

0.9.1 解复用

编码时,节目流由基本流复用而成,而传输流则由基本流、节目流或其他传输流的内容复用而成。基本流除了音频和视频位流外,还可包括私有、保留和补充位流。基本流被临时性地分割为 PES 分组包,分组包被串行化。一个 PES 分组包仅包含来自一个基本流的编码字节。

节目流中的分组包长度可以是固定的或可变的,但分组包长度应遵守 8.1 和 8.2 中的规定。传输流分组包长度是 188 字节,PES 分组包的长度可以是固定的或可变的,在大多数的应用中相对较长。

解码时需要对复用的节目流或传输流解复用,以重建基本流。这可以借助节目流分组包头中的 stream_id 和传输流分组包头中的 PEID 来完成。

0.9.2 多个基本流的同步回放

多个基本流之间的同步通过节目流或传输流中的相对显示时间(relative_display_time)来完成。时间戳通常以 90 kHz 时钟周期为单位。多个基本流解码的同步是通过调整每个流的解码至一个公共主控时钟基准,而不是通过调整一个流的解码以符合另一个流的解码。主控时钟基准可以是多个解码器时钟中的一个,也可以是数据源时钟或某个外部时钟。

一个传输流可包含多个节目,其中的每一节目都可能有自己的时钟基准。一个传输流中不同节目的时钟基准可能不同。

relative_display_time 用于单个基本流解码,它存在于传输流中相应的 TS 分组包中。编码器在捕获时记录时间戳,将该时间戳减去当前系统时间得到相对显示时间,并连同有关编码数据传输到解码器。解码器通过它们来调度节目展现,以实现端到端的同步。

一个信道中解码系统的同步是通过使用传输流中的 transport_rate 来实现的。transport_rate 是编码位流自身传输率的采样值。它们来自于同一个时间基准,该时间基准在同一个节目中也用作音频和视频的 relative_display_time 值的采样。因为每一节目可能有自己的时间基准,所以一个包含多个节目的传输流中的每个节目可有独立的 transport_rate 字段。在某些场合下,节目共享 transport_rate 字段也是可能的。确定一个节目与哪个 transport_rate 字段相关联的方法可见 6.5 节目特定信息(PSI)。一个节目有且仅有一个相关的 transport_rate 基准。在易发生错误的环境中,为了在包丢失情况下仍能正确恢复时钟,传输流分组包中可以包含包计数值 packet_count。

0.9.3 与压缩层的关系

在有些情况下,PES 分组包层是独立于压缩层的。考虑到 PES 分组包的有效负载不需像 GB/T 20090.2、GB/T 20090.3 所规定的那样以压缩层起始码开头,PES 分组包层是独立的。

示例: 视频的起始码可出现在 PES 分组包有效负载的任何位置,并且可被 PES 分组包包头所分开。

然而,在 PES 分组包包头中编码的时间戳用于决定压缩层结构(即显示单元)的显示时间。此外,当基本流数据符合 GB/T 20090.2、GB/T 20090.3 时,PES_packet_data_bytes 应遵照本部分的字节数进行字节对齐。

0.10 系统参考解码器

本部分使用了系统目标解码器(STD)来提供定时和缓冲之间关系的形式化表示。对传输流称作传输流系统目标解码器(T-STD),见 6.2;对节目流,称作节目流系统目标解码器(P-STD),见 7.2。因为 STD 是依据 GB/T 20090 中的字段(如缓冲区大小)参数化构造的,所以每一个基本流都有自己的 STD 参数。编码器产生的位流应符合适当的 STD 约束。物理解码器可以假定一个 AVS 位流能在其 STD 上正确播放。物理解码器应对其与 STD 在设计上的不同之处作出补偿。

信息技术 先进音视频编码

第 1 部分：系统

1 范围

GB/T 20090 的本部分规定了编码的系统层,主要用于组合 GB/T 20090 定义的视频位流和音频位流。系统层支持以下 5 个基本功能:

- a) 解码时多个压缩位流的同步;
- b) 多条压缩位流交织为一个单一位流;
- c) 为启动解码而对缓冲区进行初始化;
- d) 连续的缓冲区管理;
- e) 时间标识。

本部分适用于数字电视广播、激光数字存储媒体、互联网宽带流媒体、多媒体通信等应用。

一个多路复用位流可以是传输流或节目流。两种流均由 PES 分组包或包含其他必要信息的分组构成。两种流类型均支持复用来自具有一个共同时间基准节目的视频和音频压缩位流。传输流还支持复用来自具有独立时间基准的多个节目的视频和音频压缩位流。节目流通常更适用于几乎不发生差错的环境,并且支持节目信息的软件处理。传输流更适用于可能出错的环境。

无论是传输流还是节目流,一个多路复用位流的结构分为两层:最外层是系统层,最内层是压缩层。系统层提供了使用系统中一个或多个压缩数据流所需的功能。GB/T 20090 的第 2 部分和第 3 部分定义了音频和视频数据的压缩编码层。其他类型数据编码的定义不包括在本部分中,但如果它们符合本部分第 9 章的定义,则系统层也支持此类数据编码。

2 规范性引用文件

下列文件对于本文件的应用是必不可少的。凡是注日期的引用文件,仅注日期的版本适用于本文件。凡是不注日期的引用文件,其最新版本(包括所有的修改单)适用于本文件。

GB/T 4880.2—2000 语种名称代码 第 2 部分:3 字母代码(eqv ISO 639-2:1998)

GB/T 15273.1—1994 信息处理 八位单字节编码图形字符集 第 1 部分:拉丁字母一(idt ISO 8859-1:1987)

GB/T 17191.1—1997 信息技术 具有 1.5 Mbit/s 数据传输率的数字存储媒体运动图像及其伴音的编码 第 1 部分:系统(idt ISO/IEC 11172-1:1993)

GB/T 17191.2—1997 信息技术 具有 1.5 Mbit/s 数据传输率的数字存储媒体运动图像及其伴音的编码 第 2 部分:视频(idt ISO/IEC 11172-2:1993)

GB/T 17191.3—1997 信息技术 具有 1.5 Mbit/s 数据传输率的数字存储媒体运动图像及其伴音的编码 第 3 部分:音频(idt ISO/IEC 11172-3:1993)

GB/T 17975.1 信息技术 运动图像及其伴音信息的通用编码 第 1 部分:系统(GB/T 17975.1—2010,ISO/IEC 13818-1:2007,MOD)

GB/T 17975.2—2000 信息技术 运动图像及其伴音信息的通用编码 第 2 部分:视频(idt ITU-T H.262:1995)

GB/T 17975.3—2002 信息技术 运动图像及其伴音信息的通用编码 第 3 部分:音频(idt ISO/